

Dependent Randomized Rounding for Matroid Polytopes and Applications

Chandra Chekuri*

Jan Vondrák†

Rico Zenklusen‡

November 6, 2009

Abstract

Motivated by several applications, we consider the problem of randomly rounding a fractional solution in a matroid (base) polytope to an integral one. We consider the *pipage rounding* technique [5, 6, 36] and also present a new technique, *randomized swap rounding*. Our main technical results are concentration bounds for functions of random variables arising from these rounding techniques. We prove Chernoff-type concentration bounds for linear functions of random variables arising from both techniques, and also a lower-tail exponential bound for monotone submodular functions of variables arising from randomized swap rounding.

The following are examples of our applications.

- We give a $(1 - 1/e - \varepsilon)$ -approximation algorithm for the problem of maximizing a monotone submodular function subject to 1 matroid and k linear constraints, for any constant $k \geq 1$ and $\varepsilon > 0$. We also give the same result for a super-constant number k of "loose" linear constraints, where the right-hand side dominates the matrix entries by an $\Omega(\varepsilon^{-2} \log k)$ factor.
- We present a result on minimax packing problems that involve a matroid base constraint. We give an $O(\log m / \log \log m)$ -approximation for the general problem $\min\{\lambda : \exists x \in \{0, 1\}^N, x \in B(\mathcal{M}), Ax \leq \lambda b\}$ where m is the number of packing constraints. Examples include the low-congestion multi-path routing problem [34] and spanning-tree problems with capacity constraints on cuts [4, 16].
- We generalize the continuous greedy algorithm [35, 6] to problems involving multiple submodular functions, and use it to find a $(1 - 1/e - \varepsilon)$ -approximate pareto set for the problem of maximizing a constant number of monotone submodular functions subject to a matroid constraint. An example is the Submodular Welfare Problem where we are looking for an approximate pareto set with respect to individual players' utilities.

*Dept. of Computer Science, Univ. of Illinois, Urbana, IL 61801. Partially supported by NSF grant CCF-0728782. E-mail: chekuri@cs.illinois.edu

†IBM Almaden Research Center, San Jose, CA 95120. E-mail: jvondrak@us.ibm.com

‡Institute for Operations Research, ETH Zurich. E-mail: rico.zenklusen@ifor.math.ethz.ch

1 Introduction

Randomized rounding is a fundamental technique introduced by Raghavan and Thompson [29] in order to round a fractional solution of an LP into an integral solution. Numerous applications and variants have since been explored and it is a standard technique in the design of approximation algorithms and related areas. The original technique from [29] (and several subsequent papers) relies on independent rounding of the variables which allows one to use Chernoff-Hoeffding concentration bounds for linear functions of the variables; these bounds are critical for several applications in packing and covering problems. However, there are many situations in which independent rounding is not feasible due to the presence of constraints that cannot be violated by the rounded solution. Various techniques are used to handle such scenarios. To name just a few: alteration of solutions obtained by independent rounding, careful derandomization or constructive methods when probability of a feasible solution is non-zero but small (for example when using the Lovász Local Lemma), and various forms of correlated or dependent randomized rounding schemes. These methods are typically successful when one is interested in preserving the expected value of the sum of several random variables; the rounding schemes approximately preserve the expected value of each random variable and then one relies on linearity of expectation for the sum. There are, however, applications where one cannot use independent rounding and nevertheless one needs concentration bounds and/or the ability to handle non-linear objective functions such as convex or submodular functions of the variables; the work of Srinivasan [34] and others [14, 19] highlights some of these applications. Our focus in this paper is on such schemes. In particular we consider the problem of rounding a point in a *matroid polytope* to a vertex. We compare the existing approaches and propose a new rounding scheme which is simple and has multiple applications.

Background: Matroid polytopes, whose study was initiated by Edmonds in the 70's, form one of the most important classes of polytopes associated with combinatorial optimization problems. (For a definition, see Section 2.) Even though the full description of a matroid polytope is exponentially large, matroid polytopes can be optimized over, separated over, and they have strong integrality properties such as *total dual integrality*. As a consequence, the basic solution of a linear optimization problem over a matroid polytope is always integral and no rounding is necessary.

More recently, various applications emerged where a matroid constraint appears with additional constraints and/or the objective function is non-linear. In such cases, the issue of rounding a fractional solution in the matroid polytope re-appears as a non-trivial question. One such application is the submodular welfare problem [12, 22], which can be formulated as a submodular maximization problem subject to a partition matroid constraint. The rounding technique that turned out to be useful in this context is *pipage rounding* [5].

Pipage rounding was introduced by Ageev and Sviridenko [3], who used it for rounding fractional solutions in the bipartite matching polytope. They used a linear program to obtain a fractional solution to a certain problem, but the rounding procedure was based on an auxiliary (non-linear) objective. The auxiliary objective $F(x)$ was defined in such a way that $F(x)$ would always increase or stay constant throughout the rounding procedure. A comparison between $F(x)$ and the original objective yields an approximation guarantee. Calinescu et al. [5] adapted the pipage rounding technique to problems involving a matroid constraint rather than bipartite matchings. Moreover, they showed that the necessary convexity properties are satisfied whenever the auxiliary function $F(x)$ is a *multilinear extension of a submodular set function* f . This turned out to be crucial for further developments on submodular maximization problems - in particular an optimal $(1 - 1/e)$ -approximation for maximizing a monotone submodular function subject to a matroid constraint [35, 6], and a $(1 - 1/e - \varepsilon)$ -approximation for maximizing a monotone submodular function subject to a constant number of linear constraints [18]. As one of our applications, we consider a common generalization of these two problems.

Srinivasan [34], and building on his work Gandhi et al. [14], considered dependent randomized rounding for points in the bipartite matching polytope (and more generally the assignment polytope); their technique can be viewed as a randomized (and oblivious) version of pipage rounding. The motivation for this randomized scheme came from a different set of applications (see [34]). The results in [34, 14] showed *negative correlation* properties for their rounding scheme which implied concentration bounds (via [28]) that were then useful in

dealing with additional constraints. We make some observations regarding the results and applications in [3, 34, 14]. Although the schemes round a point in the assignment polytope, each constraint and objective function is restricted to depend on a subset of the edges incident to some vertex in the underlying bipartite graph. Further, several of the applications in [3, 34, 14] can be naturally modeled via a matroid constraint instead of using a bipartite graph with the above mentioned restriction; in fact the simple partition matroid suffices.

The pipage rounding technique for matroids, as presented in [5], is a deterministic procedure. However, it can be randomized similarly to Srinivasan’s work [34], and this is the variant presented in [6]. This variant starts with a fractional solution in the matroid base polytope, $y \in B(\mathcal{M})$, and produces a random base $B \in \mathcal{M}$ such that $\mathbf{E}[f(B)] \geq F(y)$; here F is the multilinear extension of the submodular function f . A further rounding stage is needed in case the starting point is inside the matroid polytope $P(\mathcal{M})$ rather than the matroid base polytope $B(\mathcal{M})$; pipage rounding has been extended to this case in [36]. In the analysis of [6, 36], the approximation guarantees are only in expectation. Stronger guarantees could be obtained and additional applications would arise if we could prove *concentration bounds* on the value of linear/submodular functions under such a rounding procedure. This is the focus of this paper.

Very recently, another application has emerged where rounding in a matroid polytope plays an essential role. Asadpour et al. [2] present a new approach to the Asymmetric Traveling Salesman problem achieving an $O(\log n / \log \log n)$ -approximation, improving upon the long-standing $O(\log n)$ -approximation. A crucial step in the algorithm is a rounding procedure, which given a fractional solution in the spanning tree polytope produces a spanning tree satisfying certain additional constraints. The authors of [2] use the technique of *maximum entropy sampling* which gives negative correlation properties and Chernoff-type concentration bounds for any linear function on the edges of the graph. Since spanning trees are bases in the *graphic matroid* for any graph, this rounding procedure also falls in the framework of randomized rounding in the matroid polytope. However, it is not clear whether the technique of [2] can be generalized to any matroid or whether it could be used in applications with a submodular objective function.

1.1 Our work

In this paper we study the problem of randomly rounding a point in a matroid polytope to a vertex of the polytope.¹ We consider the technique of *randomized pipage rounding* and also introduce a new rounding procedure called *randomized swap rounding*. Given a starting point $x \in P(\mathcal{M})$, the procedure produces a random independent set $S \in \mathcal{I}$ such that $\Pr[i \in S] = x_i$ for each element i . Our main technical results are *concentration bounds* for linear *and* submodular functions $f(S)$ under this new rounding. We demonstrate the usefulness of these concentration bounds via several applications.

The randomized swap rounding procedure bears some similarity to pipage rounding and can be used as a replacement for pipage rounding in [6, 36]. It can be also used as a replacement for maximum entropy sampling in [2]. However, it has several advantages over previous rounding procedures. It is easy to describe and implement, and it is very efficient. Moreover, thanks to the simplicity of randomized swap rounding, we are able to derive results that are not known for previous techniques. One example is the tail estimate for submodular functions, Theorem 1.4. On the other hand, our concentration bound for linear functions (Corollary 1.2) holds for a more general class of rounding techniques including pipage rounding (see also Lemma 4.1).

Randomized swap rounding starts from an arbitrary representation of a starting point $x \in P(\mathcal{M})$ as a convex combination of incidence vectors of independent sets. (This representation can be obtained by standard techniques and in some applications it is explicitly available.) Once a convex representation of the starting point is obtained, the running time of randomized swap rounding is bounded by $O(nd^2)$ calls to the membership oracle of the matroid, where d is the rank of the matroid and n is the size of the ground set. In comparison, pipage rounding performs $O(n^2)$ iterations each of which requires an expensive call to submodular function minimization (see [6]). Maximum entropy sampling for spanning trees in a graph $G = (V, E)$ is even more complicated;

¹Our results extend easily to the case of rounding a point in the polytope of an integer valued *polymatroid*. Additional applications may follow from this.

[2] does not provide an explicit running time, but it states that the procedure involves $O(|E|^2|V|\log|V|)$ iterations, where in each iteration one needs to compute a determinant (from Kirchhoff's matrix theorem) for each edge. Also, maximum entropy sampling preserves the marginal probabilities $\Pr[i \in S] = x_i$ only approximately, and the running time depends on the desired accuracy.

First, we show that randomized swap rounding as well as pipage rounding have the property that the indicator variables $X_i = [i \in S]$ have expectations exactly x_i , and are *negatively correlated*.

Theorem 1.1. *Let $(x_1, \dots, x_n) \in P(\mathcal{M})$ be a fractional solution in the matroid polytope and $(X_1, \dots, X_n) \in \{0, 1\}^n$ an integral solution obtained using either randomized swap rounding or randomized pipage rounding. Then $\mathbf{E}[X_i] = x_i$, and for any $T \subseteq [n]$, (i) $\mathbf{E}[\prod_{i \in T} X_i] \leq \prod_{i \in T} x_i$, (ii) $\mathbf{E}[\prod_{i \in T} (1 - X_i)] \leq \prod_{i \in T} (1 - x_i)$.*

This yields Chernoff-type concentration bounds for any linear function of X_1, \dots, X_n , as proved by Panconesi and Srinivasan [28] (see also Theorem 3.1 in [14]). Together with Theorem 1.1 we obtain:

Corollary 1.2. *Let $a_i \in [0, 1]$ and $X = \sum a_i X_i$, where (X_1, \dots, X_n) are obtained by either randomized swap rounding or randomized pipage rounding from a starting point $(x_1, \dots, x_n) \in P(\mathcal{M})$.*

- If $\delta \geq 0$ and $\mu \geq \mathbf{E}[X] = \sum a_i x_i$, then $\Pr[X \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$;
for $\delta \in [0, 1]$, the bound can be simplified to $\Pr[X \geq (1 + \delta)\mu] \leq e^{-\mu\delta^2/3}$.
- If $\delta \in [0, 1]$, and $\mu \leq \mathbf{E}[X] = \sum a_i x_i$, then $\Pr[X \leq (1 - \delta)\mu] \leq e^{-\mu\delta^2/2}$.

In particular, these bounds hold for $X = \sum_{i \in S} X_i$ where S is an arbitrary subset of the variables. We remark that in contrast, when randomized pipage rounding is performed on bipartite graphs, negative correlation holds only for subsets of edges incident to a fixed vertex [14].

More generally, we consider concentration properties for a monotone submodular function $f(R)$, where R is the outcome of randomized rounding. Equivalently, we can also write $f(R) = f(X_1, X_2, \dots, X_n)$ where $X_i \in \{0, 1\}$ is a random variable indicating whether $i \in S$. First, we consider a scenario where X_1, \dots, X_n are *independent* random variables. We prove that in this case, Chernoff-type bounds hold for $f(X_1, X_2, \dots, X_n)$ just like they would for a linear function.

Theorem 1.3. *Let $f : \{0, 1\}^n \rightarrow \mathbb{R}_+$ be a monotone submodular function with marginal values in $[0, 1]$. Let X_1, \dots, X_n be independent random variables in $\{0, 1\}$. Let $\mu = \mathbf{E}[f(X_1, X_2, \dots, X_n)]$. Then for any $\delta > 0$,*

- $\Pr[f(X_1, \dots, X_n) \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$.
- $\Pr[f(X_1, \dots, X_n) \leq (1 - \delta)\mu] \leq e^{-\mu\delta^2/2}$.

We remark that Theorem 1.3 can be used to simplify previous results for submodular maximization under linear constraints, where variables are rounded independently [18]. Furthermore, we prove a lower-tail bound in the dependent rounding case, where X_1, \dots, X_n are produced by randomized swap rounding.

Theorem 1.4. *Let $f(S)$ be a monotone submodular function with marginal values in $[0, 1]$, and $F(x) = \mathbf{E}[f(\hat{x})]$ its multilinear extension. Let $(x_1, \dots, x_n) \in P(\mathcal{M})$ be a point in a matroid polytope and R a random independent set obtained from it by randomized swap rounding. Let $\mu_0 = F(x_1, \dots, x_n)$ and $\delta > 0$. Then $\mathbf{E}[f(R)] \geq \mu_0$ and*

$$\Pr[f(R) \leq (1 - \delta)\mu_0] \leq e^{-\mu_0\delta^2/8}.$$

We do not know how to derive this result using only the property of negative correlations; in particular, we do not have a proof for pipage rounding, although we suspect that a similar tail estimate holds. (Weaker tail estimates involving a dependence on n follow directly from martingale concentration bounds; the main difficulty here is to obtain a bound which does not depend on n .) We remark that the tail estimate is with respect to the value of the starting point, $\mu_0 = F(x_1, \dots, x_n)$, rather than the actual expectation of $f(R)$,

which could be larger (it would be equal for a linear function f , or under independent rounding). For this reason, we do not have an upper tail bound. However, μ_0 is the value that we want to achieve in applications and hence this is the bound that we need.

Applications: We next discuss several applications of our rounding scheme. While some of the applications are concrete, others are couched in a general framework; specific instantiations lead to various applications new and old, and we defer some of these to a later version of the paper. Our rounding procedure can be used to improve the running time of some previous applications of pipage rounding [6, 36] and maximum entropy sampling [2]. In particular, our technique significantly simplifies the algorithm and analysis in the recent $O(\log n / \log \log n)$ -approximation for the Asymmetric Traveling Salesman problem [2]. In other applications, we obtain approximations with high probability instead of in expectation [6, 36]. Details of these improvements are deferred. Our new applications are as follows.

Submodular maximization subject to 1 matroid and k linear constraints. Given a monotone submodular function $f : 2^N \rightarrow \mathbb{R}_+$, a matroid \mathcal{M} on the same ground set N , and a system of k linear packing constraints $Ax \leq b$, we consider the following problem: $\max\{f(x) : x \in P(\mathcal{M}), Ax \leq b, x \in \{0, 1\}^n\}$. This problem is a common generalization of two previously studied problems, monotone submodular maximization subject to a matroid constraint [6] and subject to a constant number of linear constraints [18]. For any fixed $\varepsilon > 0$ and $k \geq 0$, we obtain a $(1 - 1/e - \varepsilon)$ -approximation for this problem, which is optimal up to the arbitrarily small ε (even for 1 matroid or 1 linear constraint [25, 11]), and generalizes the previously known results in the two special cases. We also obtain a $(1 - 1/e - \varepsilon)$ -approximation when the constraints are sufficiently "loose"; that is $b_i \geq \Omega(\varepsilon^{-2} \log k) \cdot A_{ij}$ for all i, j .

Minimax Integer Programs subject to a matroid constraint. Let \mathcal{M} be a matroid on a ground set N (let $n = |N|$). Let $B(\mathcal{M})$ be the base polytope of \mathcal{M} . We consider the problem $\min\{\lambda : Ax \leq \lambda b, x \in B(\mathcal{M}), x \in \{0, 1\}^n\}$ where $A \in \mathbb{R}_+^{m \times n}$ and $b \in \mathbb{R}_+^m$. We give an $O(\log m / \log \log m)$ -approximation for this problem, and a similar result for the min-cost version (with given packing constraints and element costs). This generalizes earlier results on minimax integer programs which were considered in the context of routing and partitioning problems [29, 23, 33, 34, 14]; the underlying matroid in these settings is the partition matroid. Another application fitting in this framework is the *minimum crossing spanning tree problem* and its geometric variant, the *minimum stabbing spanning tree problem*. We elaborate on these in Section 6.

Multiobjective optimization with submodular functions. Suppose we are given a matroid $\mathcal{M} = (N, \mathcal{I})$ and a constant number of monotone submodular functions $f_1, \dots, f_k : 2^N \rightarrow \mathbb{R}_+$. Given a set of "target values" V_1, \dots, V_k , we either find a certificate that there is no solution $S \in \mathcal{I}$ such that $f_i(S) \geq V_i$ for all i , or we find a solution S such that $f_i(S) \geq (1 - 1/e - \varepsilon)V_i$ for all i . Using the framework of multiobjective optimization [27], this implies that we can find efficiently a $(1 - 1/e - \varepsilon)$ -approximate pareto curve for the problem of maximizing k monotone submodular functions subject to a matroid constraint. A natural special case of this is the Submodular Welfare problem, where each objective function $f_i(S)$ represents the utility of player i . I.e., we can find a $(1 - 1/e - \varepsilon)$ -approximate pareto curve with respect to the utilities of the k players (for k constant). This result involves a new variant of the *continuous greedy algorithm* from [35], which in some sense optimizes multiple submodular functions at the same time. With linear objective functions f_i , we obtain the same guarantees with $1 - \varepsilon$ instead of $1 - 1/e - \varepsilon$. We give more details in Section 7.

Organization: In Section 2, we present the necessary definitions. In Section 3 the randomized swap rounding procedure is introduced. In Section 4, we prove a negative correlation property for a class of rounding procedures including randomized swap rounding and pipage rounding. In Section 5, we present our algorithm for maximizing a monotone submodular function subject to 1 matroid and k linear constraints. In Section 6, we present our results on minimax integer programs. In Section 7, we present our results on multiobjective optimization. In Appendix A, we give a complete description of randomized pipage rounding. In Appendix B, we present a generalization of swap rounding for rounding points in the matroid polytope rather than the base polytope. In Appendix C, we present our concentration bounds for submodular functions under independent rounding, and in Appendix D our lower-tail bound under randomized swap rounding.

2 Preliminaries

Matroid polytopes. Given a matroid $\mathcal{M} = (N, \mathcal{I})$ with rank function $r : 2^N \rightarrow \mathbb{Z}_+$, two polytopes associated with \mathcal{M} are the matroid polytope $P(\mathcal{M})$ and the matroid base polytope $B(\mathcal{M})$ [9] (see also [30]). $P(\mathcal{M})$ is the convex hull of characteristic vectors of the independent sets of \mathcal{M} .

$$P(\mathcal{M}) = \text{conv}\{\mathbf{1}_I : I \in \mathcal{I}\} = \{x \geq 0 : \forall S, \sum_{i \in S} x_i \leq r(S)\}$$

$B(\mathcal{M})$ is the convex hull of the characteristic vectors of the bases \mathcal{B} of \mathcal{M} , i.e. independent sets of maximum cardinality.

$$B(\mathcal{M}) = \text{conv}\{\mathbf{1}_B : B \in \mathcal{B}\} = P(\mathcal{M}) \cap \{x : \sum_{i \in N} x_i = r(N)\}.$$

Matroid exchange properties. To simplify notation, we use $+$ and $-$ for the addition and deletion of single elements from a set, for example $S - i + j$ denotes the set $(S \setminus \{i\}) \cup \{j\}$. The following base exchange property of matroids is crucial in the design of our rounding algorithm.

Theorem 2.1. *Let $\mathcal{M} = (N, \mathcal{I})$ be a matroid and let $B_1, B_2 \in \mathcal{B}$. For any $i \in B_1 \setminus B_2$ there exists $j \in B_2 \setminus B_1$ such that $B_1 - i + j \in \mathcal{B}$ and $B_2 - j + i \in \mathcal{B}$.*

To find an element j that corresponds to a given element i as described in the above theorem, one can simply check all elements in $B_2 \setminus B_1$. Thus a corresponding element j can be found by $O(d)$ calls to an independence oracle, where d is the rank of the matroid. For many matroids, a corresponding element j can be found faster. In particular, for the graphic matroid, j can be chosen to be any element $\neq i$ that lies simultaneously in the cut defined by the connected components of $B_1 - i$ and in the unique cycle in $B_2 + i$.

Submodular functions. A function $f : 2^N \rightarrow \mathbb{R}$ is submodular if for any $A, B \subseteq N$, $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$. In addition, f is monotone if $f(S) \leq f(T)$ whenever $S \subseteq T$. We denote by $f_A(i) = f(A + i) - f(A)$ the *marginal value* of i with respect to A . An important concept in recent work on submodular functions [5, 35, 6, 18, 20, 36] is the *multilinear extension* of a submodular function:

$$F(x) = \mathbf{E}[f(x)] = \sum_{S \subseteq N} f(S) \prod_{i \in S} x_i \prod_{i \in N \setminus S} (1 - x_i).$$

Rounding in the matroid polytope. A rounding procedure takes a point in the matroid polytope $x \in P(\mathcal{M})$ and rounds it to an independent set $R \in \mathcal{I}$. In its randomized version, it is oblivious to any objective function and produces a random independent set, with a distribution depending only on the starting point $x \in P(\mathcal{M})$. If the starting point is in the matroid base polytope $B(\mathcal{M})$, the rounded solution is a (random) base of \mathcal{M} .

One candidate for such a rounding procedure is *pipage rounding* [6, 36]. We give a complete description of the pipage rounding technique in the appendix. In particular, this rounding satisfies that $\Pr[i \in R] = x_i$ for each element i , and $\mathbf{E}[f(R)] \geq F(x)$ for any submodular function f and its multilinear extension F . Our new rounding, which is described in Section 3, satisfies the same properties and has additional advantages.

3 Randomized swap rounding

Let $\mathcal{M} = (N, \mathcal{I})$ be a matroid of rank $d = r(N)$ and let $n = |N|$. Randomized swap rounding is a randomized procedure that rounds a point $x \in P(\mathcal{M})$ to an independent set. We present the procedure for points in the base polytope. It can easily be generalized to round any point in the matroid polytope (see Appendix B.2).

Assume that $x \in B(\mathcal{M})$ is the point we want to round. The procedure needs a representation of x as a convex combination of bases, i.e., $x = \sum_{\ell=1}^m \beta_\ell \mathbf{1}_{B_\ell}$ with $\sum_{\ell=1}^m \beta_\ell = 1, \beta_\ell \geq 0$. Notice that by Carathéodory's

theorem there exists such a convex representation using at most n bases. In some applications, the vector x comes along with a convex representation. Otherwise, it is well-known that one can find such a convex representation in polynomial time using the fact that one can separate (or equivalently optimize) over the polytope in polynomial time (see for example [31]). For matroid polytopes, Cunningham [8] proposed a combinatorial algorithm that allows to find a convex representation of $x \in B(\mathcal{M})$ using at most n bases and whose runtime is bounded by $O(n^6)$ calls to an independence oracle. In special cases, faster algorithms are known; for example any point in the spanning tree polytope of a graph $G = (V, E)$ can be decomposed into a convex combination of spanning trees in $\tilde{O}(|V|^3|E|)$ time [13]. In general this would be the dominating term in the running time of randomized swap rounding.

Given a convex combination of bases $x = \sum_{\ell=1}^n \beta_\ell \mathbf{1}_{B_\ell}$, the procedure takes $O(nd^2)$ calls to a matroid independence oracle. The rounding proceeds in $n - 1$ stages, where in the first stage we merge the bases B_1, B_2 (randomly) into a new base C_2 , and replace $\beta_1 \mathbf{1}_{B_1} + \beta_2 \mathbf{1}_{B_2}$ in the linear combination by $(\beta_1 + \beta_2) \mathbf{1}_{C_2}$. In the k -th stage, C_k and B_{k+1} are merged into a new base C_{k+1} , and $(\sum_{\ell=1}^k \beta_\ell) \mathbf{1}_{C_k} + \beta_{k+1} \mathbf{1}_{B_{k+1}}$ is replaced in the linear combination by $(\sum_{\ell=1}^{k+1} \beta_\ell) \mathbf{1}_{C_{k+1}}$. After $n - 1$ stages, we obtain a linear combination $(\sum_{\ell=1}^n \beta_\ell) \mathbf{1}_{C_n} = \mathbf{1}_{C_n}$, and the base C_n is returned.

The procedure we use to merge two bases, called **MergeBases**, takes as input two bases B_1 and B_2 and two positive scalars β_1 and β_2 . It is described in the adjacent figure. Notice that the procedure relies heavily on the basis exchange property given by Theorem 2.1 to guarantee the existence of the elements j in the while loop. As discussed in Section 2, j can be found by checking all elements in $B_2 \setminus B_1$. Furthermore, since the cardinality of $B_1 \setminus B_2$ decreases at each iteration by one, the total number of iterations is bounded by $|B_1| = d$.

The main algorithm **SwapRound** is described in the figure. It uses **MergeBases** to repeatedly merge bases in the convex decomposition of x . For further analysis we present a different viewpoint on the algorithm, namely as a random process in the matroid base polytope. This also allows us to present the algorithm in a common framework with pipage rounding and to draw parallels between the approaches more easily.

We denote by an *elementary operation* of the swap rounding algorithm one iteration of the while loop in the **MergeBases** procedure, which is repeatedly called in **SwapRound**. Hence, an elementary operation changes two components in one of the bases used in the convex representation of the current point. For example, if the first elementary operation transforms the base B_1 into B'_1 , then this can be interpreted on the matroid base polytope as transforming the point $x = \sum_{\ell=1}^n \beta_\ell \mathbf{1}_{B_\ell}$ into $\beta_1 \mathbf{1}_{B'_1} + \sum_{\ell=2}^n \beta_\ell \mathbf{1}_{B_\ell}$. Hence, the **SwapRound** algorithm can be seen as a sequence of dn elementary operations leading to a random sequence $\mathbf{X}_0, \dots, \mathbf{X}_\tau$ where \mathbf{X}_t denotes the convex combination after t elementary operations.

4 Negative correlation for dependent rounding procedures

In this section, we prove a result which shows that the statement of Theorem 1.1 is true for a large class of random vector-valued processes that only change at most two components at a time. Theorem 1.1 then easily follows by observing that randomized swap rounding as well as pipage rounding fall in this class of random processes. The proof follows the same lines as [14] in the case of bipartite graphs. The intuitive reason for

negative correlation is that whenever a pair of variables is being modified, their sum remains constant. Hence, knowing that one variable is high can only make the expectation of another variable lower.

Lemma 4.1. *Let $\tau \in \mathbb{N}$ and let $\mathbf{X}_t = (X_{1,t}, \dots, X_{n,t})$ for $t \in \{0, \dots, \tau\}$ be a non-negative vector-valued random process with initial distribution given by $X_{i,0} = x_i$ with probability 1 $\forall i \in [n]$, and satisfying the following properties:*

1. $\mathbf{E}[\mathbf{X}_{t+1} | \mathbf{X}_t] = \mathbf{X}_t$ for $t \in \{0, \dots, \tau\}$ and $i \in [n]$.
2. \mathbf{X}_t and \mathbf{X}_{t+1} differ in at most two components for $t \in \{0, \dots, \tau - 1\}$.
3. For $t \in \{0, \dots, \tau\}$, if two components $i, j \in [n]$ change between \mathbf{X}_t and \mathbf{X}_{t+1} , then their sum is preserved: $X_{i,t+1} + X_{j,t+1} = X_{i,t} + X_{j,t}$.

Then for any $t \in \{0, \dots, \tau\}$, the components of \mathbf{X}_t satisfy $\mathbf{E}[\prod_{i \in S} X_{i,t}] \leq \prod_{i \in S} x_i \forall S \subseteq [n]$.

Proof. We are interested in the quantity $Y_t = \prod_{i \in S} X_{i,t}$. At the beginning of the process, we have $\mathbf{E}[Y_0] = \prod_{i \in S} x_i$. The main claim is that for each t , we have $\mathbf{E}[Y_{t+1} | \mathbf{X}_t] \leq Y_t$.

Let us condition on a particular configuration of variables at time t , $\mathbf{X}_t = (X_{1,t}, \dots, X_{n,t})$. We consider three cases:

- If no variable $X_i, i \in S$, is modified in step t , we have $Y_{t+1} = \prod_{i \in S} X_{i,t+1} = \prod_{i \in S} X_{i,t} = Y_t$.
- If exactly one variable $X_i, i \in S$, is modified in step t , then by property 1 of the lemma:

$$\mathbf{E}[Y_{t+1} | \mathbf{X}_t] = \mathbf{E}[X_{i,t+1} | \mathbf{X}_t] \cdot \prod_{j \in S \setminus \{i\}} X_{j,t} = \prod_{j \in S} X_{j,t} = Y_t.$$

- If two variables $X_i, X_j, i, j \in S$, are modified in step t , we use the property that their sum is preserved: $X_{i,t+1} + X_{j,t+1} = X_{i,t} + X_{j,t}$. This also implies that

$$\mathbf{E}[(X_{i,t+1} + X_{j,t+1})^2 | \mathbf{X}_t] = (X_{i,t} + X_{j,t})^2. \quad (1)$$

On the other hand, the value of each variable is preserved in expectation. Applying this to their difference, we get $\mathbf{E}[X_{i,t+1} - X_{j,t+1} | \mathbf{X}_t] = X_{i,t} - X_{j,t}$. Since $\mathbf{E}[Z^2] \geq (\mathbf{E}[Z])^2$ holds for any random variable, we get

$$\mathbf{E}[(X_{i,t+1} - X_{j,t+1})^2 | \mathbf{X}_t] \geq (X_{i,t} - X_{j,t})^2. \quad (2)$$

Combining (1) and (2), and using the formula $XY = \frac{1}{4}((X+Y)^2 - (X-Y)^2)$, we get

$$\mathbf{E}[X_{i,t+1}X_{j,t+1} | \mathbf{X}_t] \leq X_{i,t}X_{j,t}.$$

Therefore,

$$\mathbf{E}[Y_{t+1} | \mathbf{X}_t] = \mathbf{E}[X_{i,t+1}X_{j,t+1} | \mathbf{X}_t] \cdot \prod_{k \in S \setminus \{i,j\}} X_{k,t} \leq \prod_{k \in S} X_{k,t} = Y_t,$$

as claimed. By taking expectation over all configurations \mathbf{X}_t we obtain $\mathbf{E}[Y_{t+1}] \leq \mathbf{E}[Y_t]$. Consequently, $\mathbf{E}[\prod_{i \in S} X_{i,t}] = \mathbf{E}[Y_t] \leq \mathbf{E}[Y_{t-1}] \leq \dots \leq \mathbf{E}[Y_0] = \prod_{i \in S} x_i$, as claimed by the lemma. \square

Any process that satisfies the conditions of Lemma 4.1 thus also satisfies the first statement of Theorem 1.1. Furthermore, the second statement of Theorem 1.1 also follows by observing that for any process $(X_{1,t}, \dots, X_{n,t})$ that satisfies the conditions of Lemma 4.1, also the process $(1 - X_{1,t}, \dots, 1 - X_{n,t})$ satisfies the conditions. As we mentioned in Section 1, these results imply strong concentration bounds for linear functions of the variables X_1, \dots, X_n (Corollary 1.2).

Both randomized swap rounding and pipage rounding satisfy the conditions of Lemma 4.1 (proofs can be found in the Appendix). This implies Theorem 1.1. Note that the sequences \mathbf{X}_t created by randomized swap rounding or pipage rounding – besides satisfying the conditions of Lemma 4.1 – are Markovian, and hence they are vector-valued martingales.

5 Submodular maximization subject to 1 matroid and k linear constraints

In this section, we present an algorithm for the problem of maximizing a monotone submodular function subject to 1 matroid and k linear ("knapsack") constraints.

Problem definition. *Given a monotone submodular function $f : 2^N \rightarrow \mathbb{R}_+$ (by a value oracle), and a matroid $\mathcal{M} = (N, \mathcal{I})$ (by an independence oracle). For each $i \in N$, we have k parameters c_{ij} , $1 \leq j \leq k$. A set $S \subseteq N$ is feasible if $S \in \mathcal{I}$ and $\sum_{i \in S} c_{ij} \leq 1$ for each $1 \leq j \leq k$. The goal is to maximize f over all feasible sets.*

Kulik et al. gave a $(1 - 1/e - \varepsilon)$ -approximation for the same problem with a constant number of linear constraints, but *without* the matroid constraint [18]. Gupta, Nagarajan and Ravi [15] show that a knapsack constraint can in a technical sense be simulated in a black-box fashion by a collection of partition matroid constraints. Using their reduction and known results on submodular set function maximization subject to matroid constraints [12, 21], they obtain a $1/(p+q+1)$ -approximation with p knapsacks and q matroids for any $q \geq 1$ and fixed $p \geq 1$ (or $1/(p+q+\varepsilon)$ for any fixed $p \geq 1, q \geq 2$ and $\varepsilon > 0$).

5.1 Constant number of knapsack constraints

We consider first 1 matroid and a constant number k of linear constraints, in which case each linear constraint is thought of as a "knapsack" constraint. We show a $(1 - 1/e - \varepsilon)$ -approximation in this case, building upon the algorithm of Kulik, Shachnai and Tamir [18], which works for k knapsack constraints (without a matroid constraint). The basic idea is that we can add the knapsack constraints to the multilinear optimization problem

$$\max\{F(x) : x \in P(\mathcal{M})\}$$

which is used to achieve a $(1 - 1/e)$ -approximation for 1 matroid constraint [6]. Using standard techniques (partial enumeration), we get rid of all items of large value or size, and then scale down the constraints a little bit, so that we have some room for overflow in the rounding stage. We can still solve the multilinear optimization problem within a factor of $1 - 1/e$ and then round the fractional solution using randomized swap rounding (or pipage rounding). Using the fact that randomized swap rounding makes the size in each knapsack strongly concentrated, we conclude that our solution is feasible with constant probability.

Algorithm.

- Assume $0 < \varepsilon < 1/(4k^2)$. Enumerate all sets A of at most $1/\varepsilon^4$ items which form a feasible solution. (We are trying to guess the most valuable items in the optimal solution under a greedy ordering.) For each candidate set A , repeat the following.
- Let $\mathcal{M}' = \mathcal{M}/A$ be the matroid where A has been contracted. For each $1 \leq j \leq k$, let $C_j = 1 - \sum_{i \in A} c_{ij}$ be the remaining capacity in knapsack j . Let B be the set of items $i \notin A$ such that either $f_A(i) > \varepsilon^4 f(A)$ or $c_{ij} > k\varepsilon^3 C_j$ for some j (the item is relatively big compared to the size of some knapsack). Throw away all the items in B .
- We consider a reduced problem on the item set $N \setminus (A \cup B)$, with the matroid constraint \mathcal{M}' , knapsack capacities C_j , and objective function $g(S) = f_A(S)$. Define a polytope

$$P' = \left\{ x \in P(\mathcal{M}') : \forall j; \sum c_{ij} x_i \leq C_j \right\} \quad (3)$$

where $P(\mathcal{M}')$ is the matroid polytope of \mathcal{M}' . We solve (approximately) the following optimization problem:

$$\max \{G(x) : x \in (1 - \varepsilon)P'\} \quad (4)$$

where $G(x) = \mathbf{E}[g(\hat{x})]$ is the multilinear extension of $g(S)$. Since linear functions can be optimized over P' in polynomial time, we can use the continuous greedy algorithm [35] to find a fractional solution x^* within a factor of $1 - 1/e$ of optimal.

- Given a fractional solution x^* , we apply randomized pipage rounding to x^* with respect to the matroid polytope $P(\mathcal{M}')$. Call the resulting set R_A . Among all candidate sets A such that $A \cup R_A$ is feasible, return the one maximizing $f(A \cup R_A)$.

We remark that the value of this algorithm (unlike the $(1 - 1/e)$ -approximation for 1 matroid constraint) is purely theoretical, as it relies on enumeration of a huge (constant) number of elements.

Theorem 5.1. *With constant positive probability, the algorithm above returns a solution of value at least $(1 - 1/e - 3\varepsilon)OPT$.*

Proof. Consider an optimum solution O , i.e. $OPT = f(O)$. Order the elements of O greedily by decreasing marginal values, and let $A \subseteq O$ be the elements whose marginal value is at least $\varepsilon^4 OPT$. There can be at most $1/\varepsilon^4$ such elements, and so the algorithm will consider them as one of the candidate sets. We assume in the following that this is the set A chosen by the algorithm.

We consider the reduced instance, where $\mathcal{M}' = \mathcal{M}/A$ and the knapsack capacities are $C_j = 1 - \sum_{i \in A} c_{ij}$. $O \setminus A$ is a feasible solution for this instance and we have $g(O \setminus A) = f_A(O \setminus A) = OPT - f(A)$. We know that in $O \setminus A$, there are no items of marginal value more than the last item in A . In particular, $f_A(i) \leq \varepsilon^4 f(A) \leq \varepsilon^4 OPT$ for all $i \in O \setminus A$. We throw away all items where $f_A(i) > \varepsilon^4 f(A)$ but this does not affect any item in $O \setminus A$. We also throw away the set $B \subseteq N \setminus A$ of items whose size in some knapsack is more than $k\varepsilon^3 C_j$. In $O \setminus A$, there can be at most $1/(k\varepsilon^3)$ such items for each knapsack, i.e. $1/\varepsilon^3$ items in total. Since their marginal values with respect to A are bounded by $\varepsilon^4 OPT$, these items together have value $g(O \cap B) = f_A(O \cap B) \leq \varepsilon OPT$. $O' = O \setminus (A \cup B)$ is still a feasible set for the reduced problem, and using submodularity, its value is

$$g(O') = g((O \setminus A) \setminus (O \cap B)) \geq g(O \setminus A) - g(O \cap B) \geq OPT - f(A) - \varepsilon OPT.$$

Now consider the multilinear problem (4). Note that the indicator vector $\mathbf{1}_{O'}$ is feasible in P' , and hence $(1 - \varepsilon)\mathbf{1}_{O'}$ is feasible in $(1 - \varepsilon)P'$. Using the concavity of $G(x)$ along the line from the origin to $\mathbf{1}_{O'}$, we have $G((1 - \varepsilon)\mathbf{1}_{O'}) \geq (1 - \varepsilon)g(O') \geq (1 - 2\varepsilon)OPT - f(A)$. Using the continuous greedy algorithm [35], we find a fractional solution x^* of value

$$G(x^*) \geq (1 - 1/e)G((1 - \varepsilon)\mathbf{1}_{O'}) \geq (1 - 1/e - 2\varepsilon)OPT - f(A).$$

Finally, we apply randomized swap rounding (or pipage rounding) to x^* and call the resulting set R . By the construction of randomized swap rounding, R is independent in \mathcal{M}' with probability 1. However, R might violate some of the knapsack constraints.

Consider a fixed knapsack constraint, $\sum_{i \in S} c_{ij} \leq C_j$. Our fractional solution x^* satisfies $\sum c_{ij}x_i^* \leq (1 - \varepsilon)C_j$. Also, we know that all sizes in the reduced instance are bounded by $c_{ij} \leq k\varepsilon^3 C_j$. By scaling, $c'_{ij} = c_{ij}/(k\varepsilon^3 C_j)$, we can apply Corollary 1.2 with $\mu = (1 - \varepsilon)/(k\varepsilon^3)$:

$$\Pr\left[\sum_{i \in R} c_{ij} > C_j\right] \leq \Pr\left[\sum_{i \in R} c'_{ij} > (1 + \varepsilon)\mu\right] \leq e^{-\mu\varepsilon^2/3} < e^{-1/4k\varepsilon}.$$

On the other hand, consider the objective function $g(R)$. In the reduced instance, all items have value $g(i) \leq \varepsilon^4 OPT$. Let $\mu = G(x^*)/(\varepsilon^4 OPT)$. Then, Theorem 1.4 implies

$$\Pr[g(R) \leq (1 - \delta)G(x^*)] = \Pr[f(R)/(\varepsilon^4 OPT) \leq (1 - \delta)\mu] \leq e^{-\delta^2\mu/8} = e^{-\delta^2 G(x^*)/8\varepsilon^4 OPT}.$$

We set $\delta = \frac{OPT}{G(x^*)}\varepsilon$ and obtain

$$\Pr[g(R) \leq G(x^*) - \varepsilon OPT] \leq e^{-OPT/8\varepsilon^2 G(x^*)} \leq e^{-1/8\varepsilon^2}.$$

By the union bound,

$$\Pr[g(R) \leq G(x^*) - \varepsilon OPT \text{ or } \exists j; \sum_{i \in R} c_{ij} > C_j] \leq e^{-1/8\varepsilon^2} + ke^{-1/4k\varepsilon}.$$

For $\varepsilon < 1/(4k^2)$, this probability is at most $e^{-2k^4} + ke^{-k} < 1$. If this event does not occur, we have a feasible solution of value $f(R) = f(A) + g(R) \geq f(A) + G(x^*) - \varepsilon OPT \geq (1 - 1/e - 3\varepsilon)OPT$. \square

5.2 Loose packing constraints

In this section we consider the case when the number of linear packing constraints is not a fixed constant. The notation we use in this case is that of a packing integer program:

$$\max\{f(x) : x \in P(\mathcal{M}), Ax \leq b, x \in \{0,1\}^n\}.$$

Here $f : 2^N \rightarrow \mathbb{R}$ is a monotone submodular function with $n = |N|$, $\mathcal{M} = (N, \mathcal{I})$ is a matroid, $A \in \mathbb{R}_+^{k \times n}$ is a non-negative matrix and $b \in \mathbb{R}_+^k$ is a non-negative vector. This problem has been studied extensively when $f(x)$ is a linear function, in other words $f(x) = w^T x$ for some non-negative weight vector $w \in \mathbb{R}^n$. Even this case with A, b having only 0, 1 entries captures the maximum independent set problem in graphs and hence is NP-hard to approximate to within an $n^{1-\varepsilon}$ -factor for any fixed $\varepsilon > 0$. For this reason a variety of restrictions on A, b have been studied.

We consider the case when the constraints are sufficiently loose, i.e. the right-hand side b is significantly larger than entries in A : in particular, we assume $b_i \geq c \log k \cdot \max_j A_{ij}$ for $1 \leq i \leq k$. In this case, we propose a straightforward algorithm which works as follows.

Algorithm.

- Let $\varepsilon = \sqrt{6/c}$. Solve (approximately) the following optimization problem:

$$\max\{F(x) : x \in (1 - \varepsilon)P\}$$

where $F(x) = \mathbf{E}[f(\hat{x})]$ is the multilinear extension of $f(S)$, and

$$P = \{x \in P(\mathcal{M}) \mid \forall i; \sum_{j \in N} A_{ij}x_j \leq b_i\}.$$

Since linear functions can be optimized over P in polynomial time, we can use the continuous greedy algorithm [35] to find a fractional solution x^* within a factor of $1 - 1/e$ of optimal.

- Apply randomized pipage rounding to x^* with respect to the matroid polytope $P(\mathcal{M})$. If the resulting solution R satisfies the packing constraints, return R ; otherwise, fail.

Theorem 5.2. *Assume that $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$ such that $b_i \geq A_{ij}c \log k$ for all i, j and some constant $c = 6/\varepsilon^2$. Then the algorithm above gives a $(1 - 1/e - O(\varepsilon))$ -approximation with constant probability.*

We remark that it is NP-hard to achieve a better than $(1 - 1/e)$ -approximation even when $k = 1$ and the constraint is very loose ($A_{ij} = 1$ and $b_i \rightarrow \infty$) [11].

Proof. The proof is similar to that of Theorem 5.1, but simpler. We only highlight the main differences.

In the first stage we obtain a fractional solution such that $F(x^*) \geq (1-\varepsilon)(1-1/e)OPT$. Randomized swap rounding yields a random solution R which satisfies the matroid constraint. It remains to check the packing constraints. For each i , we have

$$\mathbf{E}\left[\sum_{j \in R} A_{ij}\right] = \sum_{j \in N} A_{ij}x_j^* \leq (1 - \varepsilon)b_i.$$

The variables X_j are negatively correlated and by Corollary 1.2 with $\delta = \varepsilon = \sqrt{6/c}$ and $\mu = c \log k$,

$$\Pr\left[\sum_{j \in R} A_{ij} > b_i\right] < e^{-\delta^2 \mu / 3} = \frac{1}{k^2}.$$

By the union bound, all packing constraints are satisfied with probability at least $1 - 1/k$. We assume here that $k = \omega(1)$. By using Theorem 1.4, we can also conclude that the value of the solution is at least $(1 - 1/e - O(\varepsilon))OPT$ with constant probability. \square

6 Minimax integer programs with a matroid constraint

Minimax integer programs are motivated by applications to routing and partitioning. The setup is as follows; we follow [33]. We have boolean variables $x_{i,j}$ for $i \in [p]$ and $j \in [\ell_i]$ for integers ℓ_1, \dots, ℓ_p . Let $n = \sum_{i \in [p]} \ell_i$. The goal is to minimize λ subject to:

- equality constraints: $\forall i \in [p], \sum_{j \in [\ell_i]} x_{i,j} = 1$
- a system of linear inequalities $Ax \leq \lambda \mathbf{1}$ where $A \in [0, 1]^{m \times n}$
- integrality constraints: $x_{i,j} \in \{0, 1\}$ for all i, j .

The variables $x_{i,j}, j \in [\ell_i]$ for each $i \in [p]$ capture the fact that exactly one option amongst the ℓ_i options in group i should be chosen. A canonical example is the congestion minimization problem for integral routings in graphs where for each i , the $x_{i,j}$ variables represent the different paths for routing the flow of a pair (s_i, t_i) and the matrix A encodes the capacity constraints of the edges. A natural approach is to solve the natural LP relaxation for the above problem and then apply randomized rounding by choosing independently for each i exactly one $j \in [\ell_i]$ where the probability of choosing $j \in [\ell_i]$ is exactly equal to $x_{i,j}$. This follows the randomized rounding method of Raghavan and Thompson for congestion minimization [29] and one obtains an $O(\log m / \log \log m)$ -approximation with respect to the fractional solution. Using Lovász Local Lemma (and complicated derandomization) it is possible to obtain an improved bound of $O(\log q / \log \log q)$ [23, 33] where q is the maximum number of non-zero entries in any column of A . This refined bound has various applications.

Interestingly, the above problem becomes non-trivial if we make a slight change to the equality constraints. Suppose for each $i \in [p]$ we now have an equality constraint of the form $\sum_{j \in [\ell_i]} x_{i,j} = k_i$ where k_i is an integer. For routing, this corresponds to a requirement of k_i paths for pair (s_i, t_i) . Now the standard randomized rounding doesn't quite work for this *low congestion multi-path routing problem*. Srinivasan [34], motivated by this generalized routing problem, developed dependent randomized rounding and used the negative correlation properties of this rounding to obtain an $O(\log m / \log \log m)$ -approximation. This was further generalized in [14] as randomized versions of pipage rounding in the context of other applications.

6.1 Congestion minimization under a matroid base constraint

Here we show that our dependent rounding in matroids allows a clean generalization of the type of constraints considered in several applications in [34, 14]. Let \mathcal{M} be a matroid on a ground set N . Let $B(\mathcal{M})$ be the base polytope of \mathcal{M} . We consider the problem

$$\min \left\{ \lambda : \exists x \in \{0, 1\}^N, x \in B(\mathcal{M}), Ax \leq \lambda \mathbf{1} \right\}$$

where $A \in [0, 1]^{m \times N}$. We observe that the previous problem with the variables partitioned into groups and equality constraints can be cast naturally as a special case of this matroid constraint problem; the equality constraints simply correspond to a partition matroid on the ground set of all variables $x_{i,j}$.

However, our framework is much more flexible. For example, consider the spanning tree problem with packing constraints: each edge has a weight w_e and we want to minimize the maximum load on any vertex, $\max_{v \in V} \sum_{e \in \delta(v)} w_e$. This problem also falls within our framework.

Theorem 6.1. *There is an $O(\log m / \log \log m)$ -approximation for the problem*

$$\min \{ \lambda : \exists x \in \{0, 1\}^N, x \in B(\mathcal{M}), Ax \leq \lambda \mathbf{1} \},$$

where m is the number of packing constraints, i.e. $A \in [0, 1]^{m \times N}$.

Proof. Fix a value of λ . Let $Z(\lambda) = \{j \mid \exists i; A_{ij} > \lambda\}$. We can force $x_j = 0$ for all $j \in Z(\lambda)$, because no element $j \in Z(\lambda)$ can be in a feasible solution for λ . In polynomial time, we can check the feasibility of the following LP:

$$P_\lambda = \{x \in B(\mathcal{M}) : Ax \leq \lambda \mathbf{1}, x|_{Z(\lambda)} = 0\}$$

(because we can separate over $B(\mathcal{M})$ and the additional packing constraints efficiently). By binary search, we can find (within $1 + \varepsilon$) the minimum value of λ such that $P_\lambda \neq \emptyset$. This is a lower bound on the actual optimum λ_{OPT} . We also obtain the corresponding fractional solution x^* .

We apply randomized swap rounding (or randomized pipage rounding) to x^* , obtaining a random set R . R satisfies the matroid base constraint by definition. Consider a fixed packing constraint (the i -th row of A). We have

$$\sum_{j \in N} A_{ij} x_j^* \leq \lambda$$

and all entries A_{ij} such that $x_j^* > 0$ are bounded by λ . We set $\tilde{A}_{ij} = A_{ij}/\lambda$, so that we can use Corollary 1.2. We get

$$\Pr[\sum_{j \in R} A_{ij} > (1 + \delta)\lambda] = \Pr[\sum_{j \in R} \tilde{A}_{ij} > 1 + \delta] < \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu.$$

For $\mu = 1$ and $1 + \delta = \frac{4 \log m}{\log \log m}$, this probability is bounded by

$$\Pr[\sum_{j \in R} A_{ij} > (1 + \delta)\lambda] \leq \left(\frac{e \log \log m}{4 \log m} \right)^{\frac{4 \log m}{\log \log m}} < \left(\frac{1}{\sqrt{\log m}} \right)^{\frac{4 \log m}{\log \log m}} = \frac{1}{m^2}$$

for sufficiently large m . Therefore, all m constraints are satisfied within a factor of $1 + \delta = \frac{4 \log m}{\log \log m}$ with high probability. \square

We remark that the approximation guarantee can be made an "almost additive" $O(\log m)$, in the following sense: Assuming that the optimum value is λ^* , for any fixed $\varepsilon > 0$ we can find a solution of value $\lambda \leq (1 + \varepsilon)\lambda^* + O(\frac{1}{\varepsilon} \log m)$. Scaling is important here: recall that we assumed $A \in [0, 1]^{N \times m}$. We omit the proof, which follows by a similar application of the Chernoff bound as above, with $\mu = \lambda^*$ and $\delta = \varepsilon + O(\frac{1}{\varepsilon \lambda^*} \log m)$.

Minimum Stabbing and Crossing Tree Problems: Another interesting application of Theorem 6.1, is to the minimum stabbing and crossing tree problems. Bilo et al. [4], motivated by several applications, considered the crossing spanning tree problem. The input is a graph $G = (V, E)$ and an explicit set \mathcal{C} of m cuts in G . The goal is to find a spanning tree that minimizes the number of edges crossing any cut in \mathcal{C} . The algorithm in [4] returns a tree that crosses any cut in \mathcal{C} at most $O((\log m + \log n)(\gamma^* + \log n))$ times where γ^* is the optimal solution value; the authors claim an improved bound of $O(\gamma^* \log n + \log m)$ in a subsequent version of the paper.

The minimum stabbing tree problem arises in computational geometry: the input is a set $V = \{v_1, \dots, v_n\}$ of points in \mathbb{R}^d ; it is assumed that d is a constant and the case of 2-dimensions is of particular interest. The task is to construct a spanning tree on V by connecting vertices with straight lines such that the crossing number, which is the maximum number of edges that are intersected by any hyperplane, is minimized. This problem was shown to be NP-hard by Fekete et al. [10]. It is relatively easy to see that the stabbing tree problem is a special case of the crossing spanning tree problem; the number of combinatorially distinct cuts induced by the hyperplanes is $O(n^d)$, one for each set of d points that define a hyperplane through them. Thus, the result in [4] implies that there is an algorithm for the stabbing tree problem that returns a tree with crossing number $O(\lambda^* \log n)$ where λ^* is the tree with the smallest crossing number (note that this is via the improved bound claimed by the authors of [4] in a longer version). Unaware of the work in [4], HarPeled very recently [16] gave a polynomial time algorithm for the stabbing tree problem that outputs a tree with crossing number $O(\lambda^* \log n + \log^2 n / \log \log n)$.

Both of the above problems can be cast as special cases of the minimization problem presented in Theorem 6.1, where \mathcal{M} is the graphic matroid and each row of A corresponds to the incidence vector of a cut. Theorem 6.1 implies that using dependent randomized rounding, an $O(\log n / \log \log n)$ -approximation can be obtained for the stabbing tree problem and an $O(\log m / \log \log m)$ -approximation for the crossing spanning tree problem. The approximation guarantee can be transformed into an almost additive one as well, leading to a solution of value $\lambda \leq (1 + \varepsilon)\lambda^* + O(\frac{1}{\varepsilon} \log n)$ for the stabbing tree problem and a solution of value $\gamma \leq (1 + \varepsilon)\gamma^* + O(\frac{1}{\varepsilon} \log m)$ for the crossing spanning tree problem. Note that these additive results imply a constant factor approximation if the optimal value is $\Omega(\log n)$ and $\Omega(\log m)$ respectively.

We remark that the results we obtain for the above problems can also be obtained by the maximum entropy sampling approach for spanning trees from [2]; our algorithms have the advantage of being simpler and more efficient.

6.2 Min-cost matroid bases with packing constraints

We can similarly handle the case where in addition we want to minimize a linear objective function. An example of such a problem would be a multi-path routing problem minimizing the total cost in addition to congestion. Another example is the minimum-cost spanning tree with packing constraints for the edges incident with each vertex. We remark that in case the packing constraints are simply degree bounds, strong results are known - namely, there is an algorithm that finds a spanning tree of optimal cost and violating the degree bounds by at most one [32]. In the general case of finding a matroid base satisfying certain "degree constraints", there is an algorithm [17] that finds a base of optimal cost and violating the degree constraints by an additive error of at most $\Delta - 1$, where each element participates in at most Δ constraints (e.g. $\Delta = 2$ for degree-bounded spanning trees). The algorithm of [17] also works for upper and lower bounds, violating each constraint by at most $2\Delta - 1$. See [17] for more details.

We consider a variant of this problem where the packing constraints can involve arbitrary weights and capacities. We show that we can find a matroid base of near-optimal cost which violates the packing constraints by a multiplicative factor of $O(\log m / \log \log m)$, where m is the total number of packing constraints.

Theorem 6.2. *There is a $(1 + \varepsilon, O(\log m / \log \log m))$ -bicriteria approximation for the problem*

$$\min \{c^T x : x \in \{0, 1\}^N, x \in B(\mathcal{M}), Ax \leq b\},$$

where $A \in [0, 1]^{m \times N}$ and $b \in \mathbb{R}^N$; the first guarantee is w.r.t. the cost of the solution and the second guarantee w.r.t. the overflow on the packing constraints.

Proof. We give a sketch of the proof. First, we throw away all elements that on their own violate some packing constraint. Then, we solve the following LP:

$$\min \{c^T x : x \in B(\mathcal{M}), Ax \leq b\}.$$

Let the optimum solution be x^* . We apply randomized swap rounding (or randomized pipage rounding) to x^* , yielding a random solution R . Since each of the m constraints is satisfied in expectation, and each element alone satisfies each packing constraint, we get by the same analysis as above that with high probability, R violates every constraint by a factor of $O(\log m / \log \log m)$.

Finally, the expected cost of our solution is $c^T x^* \leq OPT$. By Markov's inequality, the probability that $c(R) > (1 + \varepsilon)OPT$ is at most $1/(1 + \varepsilon) \leq 1 - \varepsilon/2$. With probability at least $\varepsilon/2 - o(1)$, $c(R) \leq (1 + \varepsilon)OPT$ and all packing constraints are satisfied within $O(\log m / \log \log m)$. \square

Let us rephrase this result in the more familiar setting of spanning trees. Given packing constraints on the edges incident with each vertex, using arbitrary weights and capacities, we can find a spanning tree of near-optimal cost, violating each packing constraint by a multiplicative factor of $O(\log m / \log \log m)$. As in the previous section, if we assume that the weights are in $[0, 1]$, this can be replaced by an additive factor of $O(\frac{1}{\varepsilon} \log m)$ while making the multiplicative factor $1 + \varepsilon$ (see the end of Section 6.1).

In the general case of matroid bases, our result is incomparable to that of [17], which provides an additive guarantee of $\Delta - 1$. (The assumption here is that each element participates in at most Δ degree constraints; in our framework, this corresponds to $A \in \{0, 1\}^{m \times N}$ with Δ -sparse columns.) When elements participate in many degree constraints ($\Delta \gg \log m$) and the degree bounds are $b_i = O(\log m)$, our result is actually stronger in terms of the packing constraint guarantee.

Asymmetric Traveling Salesman and Maximum Entropy Sampling: In a recent breakthrough, [2] obtained an $O(\log n / \log \log n)$ -approximation for the ATSP problem. A crucial ingredient in the approach is to round a point x in the spanning tree polytope to a tree T such that no cut of G contains too many edges of T , and the cost of the tree is within a constant factor of the cost of x . For this purpose, [2] uses the maximum entropy sampling approach which also enjoys negative correlation properties and hence one can get Chernoff-type bounds for linear sums of the variables; moreover T contains each edge e with probability x_e . We note that the number of cuts is exponential in n . To address this issue, [2] uses Karger's result on the number of cuts in a graph within a certain weight range: assuming that the minimum cut is at least 1, there are only $O(n^{2\alpha})$ cuts of weight in $(\alpha/2, \alpha]$ for any $\alpha \geq 1$. Maximum entropy sampling is technically quite involved and also computationally expensive. Our rounding procedures can be used in place of maximum entropy sampling to simplify the algorithm and the analysis in [2].

7 Multiobjective optimization with submodular functions

In this section, we consider the following problem: Given a matroid $\mathcal{M} = (N, \mathcal{I})$ and k monotone submodular functions $f_1, \dots, f_k : 2^N \rightarrow \mathbb{R}_+$, in what sense can we maximize $f_1(S), \dots, f_k(S)$ simultaneously over $S \in \mathcal{I}$? This question has been studied in the framework of *multiobjective optimization*, popularized in the CS community by the work of Papadimitriou and Yannakakis [27]. The set of all solutions which are optimal with respect to $f_1(S), \dots, f_k(S)$ is captured by the notion of a *pareto set*: the set of all solutions S such that for any other feasible solution S' , there exists i for which $f_i(S') < f_i(S)$. Since the pareto set in general can be exponentially large, we settle for the notion of a ε -approximate pareto set, where the condition is replaced by $f_i(S') < (1 + \varepsilon)f_i(S)$. Papadimitriou and Yannakakis show the following equivalence [27, Theorem 2]:

Proposition 7.1. *An ε -approximate pareto set can be found in polynomial time, if and only if the following problem can be solved: Given (V_1, \dots, V_k) , either return a solution with $f_i(S) \geq V_i$ for all i , or answer that there is no solution such that $f_i(S) \geq (1 + \varepsilon)V_i$ for all i .*

The latter problem is exactly what we address in this section. We show the following result.

Theorem 7.2. *For any fixed $\varepsilon > 0$ and $k \geq 2$, given a matroid $\mathcal{M} = (N, \mathcal{I})$, monotone submodular functions $f_1, \dots, f_k : 2^N \rightarrow \mathbb{R}_+$, and values $V_1, \dots, V_k \in \mathbb{R}_+$, in polynomial time we can either*

- find a solution $S \in \mathcal{I}$ such that $f_i(S) \geq (1 - 1/e - \varepsilon)V_i$ for all i , or
- return a certificate that there is no solution with $f_i(S) \geq V_i$ for all i .

If $f_i(S)$ are linear functions, the guarantee in the first case becomes $f_i(S) \geq (1 - \varepsilon)V_i$.

This together with Proposition 7.1 implies that for any constant number of linear objective functions subject to a matroid constraint, an ε -approximate pareto set can be found in polynomial time. (This was known in the case of multiobjective spanning trees [27].) Furthermore, a straightforward modification of Prop. 7.1 (see [27], Theorem 2) implies that for monotone submodular functions $f_i(S)$, we can find a $(1 - 1/e - \varepsilon)$ -approximate pareto set.

Our algorithm requires a modification of the continuous greedy algorithm from [35, 6]. We show the following, which might be useful in other applications as well. In the following lemma, we do not require k to be constant.

Lemma 7.3. Consider monotone submodular functions $f_1, \dots, f_k : 2^N \rightarrow \mathbb{R}_+$, their multilinear extensions $F_i(x) = \mathbf{E}[f_i(\hat{x})]$ and a down-monotone polytope $P \subset \mathbb{R}_+^N$ such that we can optimize linear functions over P in polynomial time. Then given $V_1, \dots, V_k \in \mathbb{R}_+$ we can either

- find a point $x \in P$ such that $F_i(x) \geq (1 - 1/e)V_i$ for all i , or
- return a certificate that there is no point $x \in P$ such that $F_i(x) \geq V_i$ for all i .

Proof. We refer to Section 2.3 of [6] for intuition and notation. Assuming that there is a solution $S \in \mathcal{I}$ achieving $f_i(S) \geq V_i$, Section 2.3 in [6] implies that for any fractional solution $y \in P(\mathcal{M})$ there is a direction $v^*(y) \in P(\mathcal{M})$ such that $v^*(y) \cdot \nabla F_i(y) \geq V_i - F_i(y)$. Moreover, the way this direction is constructed is by going towards the actual optimum - i.e., this direction is the same for all i . Assuming that such a direction exists, we can find it by linear programming. If the LP is infeasible, we have a certificate that there is no solution satisfying $f_i(S) \geq V_i$ for all i . Otherwise, we follow the continuous greedy algorithm and the analysis implies that

$$\frac{dF_i}{dt} \geq v^*(y(t)) \cdot \nabla F_i(y(t)) \geq V_i - F_i(y(t))$$

which implies $F_i(y(1)) \geq (1 - 1/e)V_i$. □

Given Lemma 7.3, we sketch the proof of Theorem 7.2 as follows. First, we guess a constant number of elements so that for each remaining element j , the marginal value for each i is at most $\varepsilon^3 V_i$. In the following, we just assume that $f_i(j) \leq \varepsilon^3 V_i$ for all i, j . For each objective function f_i , we consider the multilinear relaxation of the problem:

$$\max\{F_i(x) : x \in P(\mathcal{M})\}$$

where $F_i(x) = \mathbf{E}[f_i(\hat{x})]$. We apply Lemma 7.3 to find a fractional solution y^* satisfying $F_i(y^*) \geq (1 - 1/e)V_i$ for all i (or a certificate that there is no solution $y \in P(\mathcal{M})$ such that $F_i(y) \geq V_i$ for all i ; this implies that there is no feasible solution S such that $f_i(S) \geq V_i$ for all i). For linear objective functions, the problem is much simpler: then $F_i(x)$ are linear functions and we can find a fractional solution satisfying $F_i(y^*) \geq V_i$ directly by linear programming.

We apply randomized swap rounding to y^* , to obtain a random solution $R \in \mathcal{I}$ satisfying the lower-tail concentration bound of Theorem 1.4. The marginal values of f_i are bounded by $\varepsilon^3 V_i$, so by standard scaling we obtain

$$\Pr[f_i(R) < (1 - \delta)F_i(y^*)] < e^{-\delta^2 F_i(y^*) / 8\varepsilon^3 V_i} \leq e^{-\delta^2 / 16\varepsilon^3}.$$

Hence, we can set $\delta = \varepsilon$ and obtain error probability at most $e^{-1/16\varepsilon}$. By the union bound, the probability that $f_i(R) < (1 - \varepsilon)F_i(y^*)$ for any i is at most $ke^{-1/16\varepsilon}$. For sufficiently small $\varepsilon > 0$, this is a constant probability smaller than 1. Then, $f_i(R) \geq (1 - 1/e - \varepsilon)V_i$ for all i . This proves Theorem 7.2.

To conclude, we are able to find a $(1 - 1/e - \varepsilon)$ -approximate pareto set for any constant number of monotone submodular functions and any matroid constraint. This has a natural interpretation in the setting of the Submodular Welfare Problem (which is a special case, see [12, 22]). Then each objective function $f_i(S)$ is the utility function of a player, and we want to find a pareto set with respect to all possible allocations. To summarize, we can find a set of all allocations that are not dominated by any other allocation within a factor of $1 - 1/e - \varepsilon$ per player.

Acknowledgments: CC and JV thank Anupam Gupta for asking about the approximability of maximizing a monotone submodular set function subject to a matroid constraint and a constant number of knapsack constraints; this motivated their work. JV thanks Ilias Diakonikolas for fruitful discussions concerning multiobjective optimization that inspired the application in Section 7. RZ is grateful to Michel Goemans for introducing him to a version of Shannon’s switching game that inspired the randomized swap rounding algorithm. We thank Mohit Singh for pointing out [4, 16].

References

- [1] N. Alon and J. Spencer. *The Probabilistic Method*. 2nd Edition, John Wiley & Sons.
- [2] A. Asadpour, M. Goemans, A. Madry, S.O. Gharan, and A. Saberi. An $O(\log n / \log \log n)$ -approximation algorithm for the assymetric traveling salesman problem. To appear in *Proc. of ACM-SIAM SODA*, 2010.
- [3] A. Ageev and M. Sviridenko. Pipage rounding: a new method of constructing algorithms with proven performance guarantee. *J. of Combinatorial Optimization*, 8:307–328, 2004.
- [4] V. Bilo, V. Goyal, R. Ravi and M. Singh. On the Crossing Spanning Tree Problem. *Proc. of APPROX*, 51–60, 2004.
- [5] G. Calinescu, C. Chekuri, M. Pál and J. Vondrák. Maximizing a submodular set function subject to a matroid constraint. *Proc. of IPCO*, 182–196, 2007.
- [6] G. Calinescu, C. Chekuri, M. Pál and J. Vondrák. Maximizing a submodular set function subject to a matroid constraint. To appear in *SIAM Journal on Computing, special issue for STOC 2008*.
- [7] P.M. Camerini, G. Galbiati, and F. Maffioli. Random pseudo-polynomial algorithms for exact matroid problems. *J. of algorithms* 13, 258–273, 1992.
- [8] W. H. Cunningham. Testing membership in matroid polyhedra. *Journal of Combinatorial Theory, Series B*, 36(2):161–188, April 1984.
- [9] J. Edmonds. Matroids, submodular functions and certain polyhedra. *Combinatorial Structures and Their Applications*, 69–87, 1970.
- [10] S. P. Fekete, M. E. Lübbecke and H. Meijer. Minimizing the stabbing number of matchings, trees, and triangulations. *Proc. of ACM-SIAM SODA*, 437–446, 2004.
- [11] U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.
- [12] M. L. Fisher, G. L. Nemhauser and L. A. Wolsey. An analysis of approximations for maximizing submodular set functions - II. *Math. Prog. Study*, 8:73–87, 1978.
- [13] H.N. Gabow and K.S. Manu. Packing algorithms for arborescences (and spanning trees) in capacitated graphs. *Mathematical Programming* 82(1–2):83–109, 1998.

- [14] R. Gandhi, S. Khuller, S. Parthasarathy and A. Srinivasan. Dependent rounding and its applications to approximation algorithms. *Journal of the ACM* 53:324–360, 2006.
- [15] A. Gupta, V. Nagarajan and R. Ravi. Personal communication, 2009.
- [16] S. Har-Peled. Approximating spanning trees with low crossing number. Technical report, <http://valis.cs.uiuc.edu/~sariel/papers/09/crossing/>.
- [17] T. Király, L. C. Lau, and M. Singh. Degree bounded matroids and submodular flows. *Proc. of IPCO*, 2008.
- [18] A. Kulik, H. Shachnai and T. Tamir. Maximizing submodular set functions subject to multiple linear constraints. *Proc. of ACM-SIAM SODA*, 545–554, 2009.
- [19] V. S. A. Kumar, M. V. Marathe, S. Parthasarathy and A. Srinivasan. A Unified Approach to Scheduling on Unrelated Parallel Machines. *JACM*, Vol. 56, 2009.
- [20] J. Lee, V. Mirrokni, V. Nagarajan and M. Sviridenko. Maximizing non-monotone submodular functions under matroid and knapsack constraints. *Proc. of ACM STOC 2009*, 323–332.
- [21] J. Lee, M. Sviridenko, and J. Vondrák. Submodular maximization over multiple matroids via generalized exchange properties. *Proc. of APPROX*, Springer LNCS, 244–257, 2009.
- [22] B. Lehmann, D. J. Lehmann, and N. Nisan. Combinatorial auctions with decreasing marginal utilities. *Games and Economic Behavior* 55:270–296, 2006.
- [23] T. Leighton, C.-J. Lu, S. Rao, and A. Srinivasan. New algorithmic aspects of the local lemma with applications to routing and partitioning. *SIAM J. on Computing*, Vol. 31, 626–641, 2001.
- [24] G. L. Nemhauser, L. A. Wolsey and M. L. Fisher. An analysis of approximations for maximizing submodular set functions - I. *Math. Prog.*, 14:265–294, 1978.
- [25] G. L. Nemhauser and L. A. Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Math. Oper. Research*, 3(3):177–188, 1978.
- [26] J. Pach and P. K. Agarwal. Combinatorial Geometry Wiley-Interscience, 1995.
- [27] C. H. Papadimitriou and M. Yannakakis. The complexity of tradeoffs, and optimal access of web sources. In *Proc. of FOCS 2000*, 86–92.
- [28] A. Panconesi and A. Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM Journal on Computing* 26:350–368, 1997.
- [29] P. Raghavan and C. D. Thompson. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica* 7(4):365–374, 1987.
- [30] A. Schrijver. Combinatorial optimization - polyhedra and efficiency. Springer, 2003.
- [31] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.
- [32] M. Singh and L.C. Lau. Approximating minimum bounded degree spanning trees to within one of optimal, *Proc. of ACM STOC* (2007).
- [33] A. Srinivasan. An extension of the Lovász Local Lemma, and its applications to integer programming. *SIAM J. on Computing*, Vol 36, 609–634, 2006. Preliminary version in *Proc. of ACM-SIAM SODA*, 1996.

- [34] A. Srinivasan. Distributions on level-sets with applications to approximation algorithms, *Proc. IEEE Symposium on Foundations of Computer Science (FOCS)*, 588–597, 2001.
- [35] J. Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. *Proc. of ACM STOC*, 67–74, 2008.
- [36] J. Vondrák. Symmetry and approximability of submodular maximization problems. In *Proc. of IEEE FOCS*, 251–270, 2009.

A Randomized pipage rounding

Let us summarize the pipage rounding technique in the context of matroid polytopes [5, 6]. The basic version of the technique assumes that we start with a point in the matroid base polytope, and we want to round it to a vertex of $B(\mathcal{M})$. In each step, we have a fractional solution $y \in B(\mathcal{M})$ and a *tight set* T (satisfying $y(T) = r(T)$) containing at least two fractional variables. We modify the two fractional variables in such a way that their sum remains constant, until some variable becomes integral or a new constraint becomes tight. If a new constraint becomes tight, we continue with a new tight set, which can be shown to be a proper subset of the previous tight set [5, 6]. Hence, after n steps we produce a new integral variable, and the process terminates after n^2 steps.

In the randomized version of the technique, each step is randomized in such a way that the expectation of each variable is preserved. Here is the randomized version of pipage rounding [6]. The subroutine **HitConstraint**(y, i, j) starts from y and tries to increase y_i and decrease y_j at the same rate, as long as the the solution is inside $B(\mathcal{M})$. It returns a new point y and a tight set A , which would be violated if we go any further. This is used in the main algorithm **PipageRound**(\mathcal{M}, y), which repeats the process until an integral solution in $B(\mathcal{M})$ is found.

Subroutine HitConstraint(y, i, j):

```

Denote  $\mathcal{A} = \{A \subseteq X : i \in A, j \notin A\}$ ;
Find  $\delta = \min_{A \in \mathcal{A}}(r_{\mathcal{M}}(A) - y(A))$ 
and a set  $A \in \mathcal{A}$  attaining the above minimum;
If  $y_j < \delta$  then  $\{\delta \leftarrow y_j, A \leftarrow \{j\}\}$ ;
 $y_i \leftarrow y_i + \delta, y_j \leftarrow y_j - \delta$ ;
Return  $(y, A)$ .
```

Algorithm PipageRound((\mathcal{M}, y)):

```

While ( $y$  is not integral) do
   $T \leftarrow X$ ;
  While ( $T$  contains fractional variables) do
    Pick  $i, j \in T$  fractional;
     $(y^+, A^+) \leftarrow \text{HitConstraint}(y, i, j)$ ;
     $(y^-, A^-) \leftarrow \text{HitConstraint}(y, j, i)$ ;
     $p \leftarrow \|y^+ - y\| / \|y^+ - y^-\|$ ;
    With probability  $p$ ,  $\{y \leftarrow y^-, T \leftarrow T \cap A^-\}$ ;
    Else  $\{y \leftarrow y^+, T \leftarrow T \cap A^+\}$ ;
  EndWhile
EndWhile
Output  $y$ .
```

Subsequently [36], pipage rounding was extended to the case when the starting point is in the matroid polytope $P(\mathcal{M})$, rather than $B(\mathcal{M})$. This is not an issue in [6], but it is necessary for applications with non-monotone submodular functions [36] or with additional constraints, such as in this paper.

The following procedure takes care of the case when we start with a fractional solution $x \in P(\mathcal{M})$. It adjusts the solution in a randomized way so that the expectation of each variable is preserved, and the new fractional solution is in the base polytope of a (possibly reduced) matroid.

*Algorithm **Adjust**((\mathcal{M}, x)):*

```

While ( $x$  is not in  $B(\mathcal{M})$ ) do
    If (there is  $i$  and  $\delta > 0$  such that  $x + \delta \mathbf{e}_i \in P(\mathcal{M})$ ) do
        Let  $x_{max} = x_i + \max\{\delta : x + \delta \mathbf{e}_i \in P(\mathcal{M})\}$ ;
        Let  $p = x_i/x_{max}$ ;
        With probability  $p$ ,  $\{x_i \leftarrow x_{max}\}$ ;
        Else  $\{x_i \leftarrow 0\}$ ;
    EndIf
    If (there is  $i$  such that  $x_i = 0$ ) do
        Delete  $i$  from  $\mathcal{M}$  and remove the  $i$ -coordinate from  $x$ .
    EndIf
EndWhile
Output ( $\mathcal{M}, x$ ).
```

To summarize, the complete procedure works as follows. For a given $x \in P(\mathcal{M})$, we run $(\mathcal{M}', y) := \text{Adjust}(\mathcal{M}, x)$, followed by **PipageRound**((\mathcal{M}', y)). The outcome is a base in the restricted matroid where some elements have been deleted, i.e. an independent set in the original matroid.

B Proofs and generalizations for randomized swap rounding

In this section we proof that randomized swap rounding satisfies the conditions of Lemma 4.1 and generalize the procedure to points in the matroid polytope.

B.1 Proof of conditions for negative correlation

Lemma B.1. *Randomized swap rounding satisfies the conditions of Lemma 4.1.*

Proof. Let $X_{i,t}$ denote the i -th component of \mathbf{X}_t . To prove the first condition of Lemma 4.1 we condition on a particular vector \mathbf{X}_t at time t of the process and on its convex representation $\mathbf{X}_t = \sum_{\ell=1}^k \beta_\ell \mathbf{1}_{B_\ell}$. The vector \mathbf{X}_{t+1} is obtained from \mathbf{X}_t by an elementary operation. Without loss of generality we assume that the elementary operation does a swap between the bases B_1 and B_2 involving the elements $i \in B_1 \setminus B_2$ and $j \in B_2 \setminus B_1$. Let B'_1 and B'_2 be the bases after the swap. Hence, with probability $\beta_1/(\beta_1 + \beta_2)$, $B'_1 = B_1$ and $B'_2 = B_2 - j + i$, and with probability $\beta_2/(\beta_1 + \beta_2)$, $B'_1 = B_1 - i + j$ and $B'_2 = B_2$. Thus,

$$\begin{aligned} \mathbf{E}[\beta_1 \mathbf{1}_{B'_1} + \beta_2 \mathbf{1}_{B'_2}] &= \frac{\beta_1}{\beta_1 + \beta_2} (\beta_1 \mathbf{1}_{B_1} + \beta_2 (\mathbf{1}_{B_2} - \mathbf{e}_j + \mathbf{e}_i)) + \frac{\beta_2}{\beta_1 + \beta_2} (\beta_1 (\mathbf{1}_{B_1} - \mathbf{e}_i + \mathbf{e}_j) + \beta_2 \mathbf{1}_{B_2}) \\ &= \beta_1 \mathbf{1}_{B_1} + \beta_2 \mathbf{1}_{B_2}, \end{aligned}$$

where $\mathbf{e}_i = \mathbf{1}_{\{i\}}$ and $\mathbf{e}_j = \mathbf{1}_{\{j\}}$ denote the canonical basis vectors corresponding to element i and j , respectively. Since the vector \mathbf{X}_{t+1} is given by $\mathbf{X}_{t+1} = \beta_1 \mathbf{1}_{B'_1} + \beta_2 \mathbf{1}_{B'_2} + \sum_{\ell=3}^k \beta_\ell \mathbf{1}_{B_\ell}$, we obtain $\mathbf{E}[\mathbf{X}_{t+1} | \mathbf{X}_t] = \mathbf{X}_t$. The second condition of Lemma 4.1 is satisfied since an elementary operation only changes two elements in one base of the convex representation as discussed above. To check the third condition of the lemma, assume without loss of generality that \mathbf{X}_{t+1} is obtained from $\mathbf{X}_t = \sum_{\ell=1}^k \beta_\ell \mathbf{1}_{B_\ell}$ by replacing B_1 by $B_1 - i + j$. Hence, $X_{i,t+1} = X_{i,t} + \beta_1$ and $X_{j,t+1} = X_{j,t} - \beta_1$, implying that the third condition of the lemma is satisfied. \square

B.2 Adapting randomized swap rounding to points in the matroid polytope

In this section we show how randomized swap rounding can be generalized to round a point in the matroid polytope to an independent set, such that the conditions of Lemma 4.1 are still satisfied. We first present a generalization where the rounding is done by applying randomized swap rounding for base polytopes to an extension of the underlying matroid. In a second step we show that this procedure can easily be interpreted as a procedure on the initial matroid, leading to a simpler description of the method. An advantage of presenting the method as a special case of base rounding, is that results presented for randomized swap rounding on base polytopes easily carry over to the general rounding procedure.

Let $x \in P(\mathcal{M})$ be the point to round. Similar as for the base polytope case, we need a representation of x as a convex combination of independent sets. Again, the algorithm of Cunningham [8] can be used to obtain a convex combination of x using at most $n + 1$ independent sets with a running time which is bounded by $O(n^6)$ oracle calls. Thus, we assume that such a convex combination of x using $n + 1$ independent sets $I_1, \dots, I_{n+1} \in \mathcal{I}$ is given, i.e., $x = \sum_{\ell=1}^{n+1} \beta_\ell \mathbf{1}_{I_\ell}$.

Let $\mathcal{M}' = (N', \mathcal{I}')$ be the following extension of the matroid $\mathcal{M} = (N, \mathcal{I})$. The set N' is obtained from N by adding d additional dummy elements $\{s_1, \dots, s_d\}$, $N' = N \cup \{s_1, \dots, s_d\}$. The independent sets are defined by $\mathcal{I}' = \{I \subseteq N' \mid I \cap N \in \mathcal{I}, |I| \leq d\}$. Thus, a base of \mathcal{M} is also a base of \mathcal{M}' . The task of rounding x in \mathcal{M} can be transformed into rounding a point in the base polytope of \mathcal{M}' as follows. Every independent set I_ℓ that is used in the convex representation of x , is extended to a base B'_ℓ of \mathcal{M}' by adding an arbitrary subset of $\{s_1, \dots, s_d\}$ of cardinality $d - |I_\ell|$. Hence, $y = \sum_{\ell=1}^{n+1} \beta_\ell \mathbf{1}_{B'_\ell}$ is a point in the base polytope of \mathcal{M}' . Then the randomized swap rounding procedure as presented in Section 3 for points in the base polytope is used to get a point $\mathbf{1}_{B'}$ in $B(\mathcal{M}')$. The point $\mathbf{1}_{B'}$ is finally transformed into a point \bar{x} that is a vertex of $P(\mathcal{M})$ by projecting $\mathbf{1}_{B'}$ onto the components corresponding to elements in N . The point \bar{x} is returned by the algorithm. By Lemma B.1, the random point $\mathbf{1}_{B'}$ satisfies the conditions of Lemma 4.1. Since the projection does not change the distribution of the components of $\mathbf{1}_{B'}$, also \bar{x} satisfies the same properties.

The dummy elements can be interpreted as elements that do not have any influence in the final outcome, since they will be removed by the projection. Consider for example an elementary operation on two bases $B'_1, B'_2 \in \mathcal{B}$ which are extensions of two independent sets $I_1, I_2 \in \mathcal{I}$ to the matroid \mathcal{M}' , and let $i \in B'_1 \setminus B'_2$ and $j \in B'_2 \setminus B'_1$ be the two elements involved in the swap. If i is a dummy element, i.e., $i \in \{s_1, \dots, s_d\}$, then replacing B'_2 by $B'_2 - j + i$ corresponds to removing element j from I_2 .

Consider the above algorithm using dummy elements with the following modification: At each elementary operation, if possible, two non-dummy elements are chosen. One can easily observe that describing this version of the algorithm without dummy elements corresponds to replacing the **MergeBases** procedure with the following procedure to merge two independent sets. The procedure, called **MergeIndepSets**, takes two independent sets $I_1, I_2 \in \mathcal{I}$ and two positive scalars β_1, β_2 as input. To simplify the description of the procedure, we assume $|I_1| \geq |I_2|$, otherwise the roles of I_1 and I_2 have to be exchanged in the algorithm.

Algorithm MergeIndepSets($\beta_1, I_1, \beta_2, I_2$):

```

Find a set  $S \subseteq I_1 \setminus I_2$  of cardinality  $|I_1| - |I_2|$  such that  $I_2 \cup S \in \mathcal{I}$ ;
 $I'_2 = I_2 \cup S$ ;
While ( $I_1 \neq I'_2$ ) do
    Pick  $i \in I_1 \setminus I'_2$  and find  $j \in I'_2 \setminus I_1$  such that  $I_1 - i + j \in \mathcal{I}$  and  $I'_2 - j + i \in \mathcal{I}$ ;
    With probability  $\beta_1/(\beta_1 + \beta_2)$ ,  $\{I'_2 \leftarrow I'_2 - j + i\}$ ;
    Else  $\{I_1 \leftarrow I_1 - i + j\}$ ;
EndWhile
For ( $i \in S$ ) do
    With probability  $\beta_2/(\beta_1 + \beta_2)$ ,  $\{I_1 \leftarrow I_1 - i\}$ ;
EndFor
Output  $I_1$ .

```

The existence of a set S as used in the algorithm easily follows from the matroid axioms [30]. It can be found

by successively choosing elements in $I_1 \setminus I_2$ that can be added to I_2 still maintaining independence. Once the element $i \in I_1 \setminus I'_2$ is chosen in the while loop of the algorithm, the existence of an element $j \in I'_2 \setminus I_1$ satisfying $I_1 - i + j \in \mathcal{I}$ and $I'_2 - j + i \in \mathcal{I}$ is guaranteed by applying Theorem 2.1 to the matroid $\mathcal{M}' = (N, \mathcal{I}')$ given by $\mathcal{I}' = \{I \in \mathcal{I} \mid |I| \leq |I_1|\}$.

C Chernoff bounds for submodular functions

Here we prove Theorem 1.3, a Chernoff-type bound for a monotone submodular function $f(X_1, \dots, X_n)$ where $X_1, \dots, X_n \in \{0, 1\}$ are independent random variables. Similarly to the proof of Chernoff bounds for linear functions, the main trick is to prove a bound on the exponential moments $\mathbf{E}[e^{\lambda f(X_1, \dots, X_n)}]$. For that purpose, we write the value of $f(X_1, \dots, X_n)$ as follows: $f(X_1, \dots, X_n) = \sum_{i=1}^n Y_i$, where

$$Y_i = f(X_1, \dots, X_i, 0, \dots, 0) - f(X_1, \dots, X_{i-1}, 0, \dots, 0).$$

The new complication is that the variables Y_i are not independent. There could be negative and even positive correlations between Y_i, Y_j . What is important for us, however, is that we can show negative correlation between $e^{\lambda \sum_{i=1}^{k-1} Y_i}$ and $e^{\lambda Y_k}$, and by induction the following bound.

Lemma C.1. *For any $\lambda \in \mathbb{R}$, a monotone submodular function and Y_1, \dots, Y_n defined as above,*

$$\mathbf{E}[e^{\lambda \sum_{i=1}^n Y_i}] \leq \prod_{i=1}^n \mathbf{E}[e^{\lambda Y_i}].$$

Proof. Denote $p_i = \Pr[X_i = 1]$. For any k , we have

$$\begin{aligned} \mathbf{E}[e^{\lambda \sum_{i=1}^k Y_i}] &= \mathbf{E}[e^{\lambda f(X_1, \dots, X_k, 0, \dots, 0)}] \\ &= p_k \mathbf{E}[e^{\lambda f(X_1, \dots, X_{k-1}, 1, \dots, 0)}] + (1 - p_k) \mathbf{E}[e^{\lambda f(X_1, \dots, X_{k-1}, 0, \dots, 0)}] \\ &= p_k \mathbf{E}[e^{\lambda f(X_1, \dots, X_{k-1}, 0, \dots, 0)} e^{\lambda F_k(X_1, \dots, X_{k-1}, 0, \dots, 0)}] + (1 - p_k) \mathbf{E}[e^{\lambda f(X_1, \dots, X_{k-1}, 0, \dots, 0)}] \end{aligned}$$

where

$$F_k(X_1, \dots, X_{k-1}, 0, \dots, 0) = f(X_1, \dots, X_{k-1}, 1, \dots, 0) - f(X_1, \dots, X_{k-1}, 0, \dots, 0)$$

denotes the marginal value of X_k being set to 1, given the preceding variables. Observe that $\mathbf{E}[F_k(X_1, \dots, X_{k-1}, 0, \dots, 0)] = \mathbf{E}[Y_k \mid X_k = 1]$.

By submodularity, F_k is a decreasing function of (X_1, \dots, X_{k-1}) . On the other hand, $\sum_{i=1}^{k-1} Y_i = f(X_1, \dots, X_{k-1}, 0, \dots, 0)$ is an increasing function of (X_1, \dots, X_{k-1}) . We get the same monotonicity properties for the exponential functions $e^{\lambda f(\dots)}$ and $e^{\lambda F_k(\dots)}$ (with a switch in monotonicity for $\lambda < 0$). By the FKG inequality, $e^{\lambda f(X_1, \dots, X_{k-1}, 0, \dots, 0)}$ and $e^{\lambda F_k(X_1, \dots, X_{k-1}, 0, \dots, 0)}$ are negatively correlated, and we get

$$\begin{aligned} \mathbf{E}[e^{\lambda f(X_1, \dots, X_{k-1}, 0, \dots, 0)} e^{\lambda F_k(X_1, \dots, X_{k-1}, 0, \dots, 0)}] &\leq \mathbf{E}[e^{\lambda f(X_1, \dots, X_{k-1}, 0, \dots, 0)}] \mathbf{E}[e^{\lambda F_k(X_1, \dots, X_{k-1}, 0, \dots, 0)}] \\ &= \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \mathbf{E}[e^{\lambda Y_k} \mid X_k = 1]. \end{aligned}$$

Hence, we have

$$\begin{aligned} \mathbf{E}[e^{\lambda \sum_{i=1}^k Y_i}] &\leq p_k \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \mathbf{E}[e^{\lambda Y_k} \mid X_k = 1] + (1 - p_k) \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \\ &= \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \cdot (p_k \mathbf{E}[e^{\lambda Y_k} \mid X_k = 1] + (1 - p_k) \cdot 1) \\ &= \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \cdot (p_k \mathbf{E}[e^{\lambda Y_k} \mid X_k = 1] + (1 - p_k) \mathbf{E}[e^{\lambda Y_k} \mid X_k = 0]) \\ &= \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \cdot \mathbf{E}[e^{\lambda Y_k}]. \end{aligned}$$

By induction, we obtain the lemma. \square

Given this lemma, we can finish the proof of Theorem 1.3 following the same outline as of proof of the Chernoff bound.

Proof. Let $Y_i = f(X_1, \dots, X_k, 0, \dots, 0) - f(X_1, \dots, X_{k-1}, 0, \dots, 0)$ as above. Let us denote $\mathbf{E}[Y_i] = \omega_i$ and $\mu = \sum_{i=1}^n \omega_i = \mathbf{E}[f(X_1, \dots, X_n)]$. By the convexity of the exponential and the fact that $Y_i \in [0, 1]$,

$$\mathbf{E}[e^{\lambda Y_i}] \leq \omega_i e^\lambda + (1 - \omega_i) = 1 + (e^\lambda - 1)\omega_i \leq e^{(e^\lambda - 1)\omega_i}.$$

Lemma C.1 then implies

$$\mathbf{E}[e^{\lambda f(X_1, \dots, X_n)}] = \mathbf{E}[e^{\lambda \sum_{i=1}^n Y_i}] \leq \prod_{i=1}^n \mathbf{E}[e^{\lambda Y_i}] \leq e^{(e^\lambda - 1)\mu}.$$

For the upper-tail bound, we use Markov's inequality as follows:

$$\Pr[f(X_1, \dots, X_n) \geq (1 + \delta)\mu] = \Pr[e^{\lambda f(X_1, \dots, X_n)} \geq e^{\lambda(1 + \delta)\mu}] \leq \frac{\mathbf{E}[e^{\lambda f(X_1, \dots, X_n)}]}{e^{\lambda(1 + \delta)\mu}} \leq \frac{e^{(e^\lambda - 1)\mu}}{e^{\lambda(1 + \delta)\mu}}.$$

We choose $e^\lambda = 1 + \delta$ which yields

$$\Pr[f(X_1, \dots, X_n) \geq (1 + \delta)\mu] \leq \frac{e^{\delta\mu}}{(1 + \delta)^{(1 + \delta)\mu}}.$$

For the lower-tail bound, we use Markov's inequality with $\lambda < 0$ as follows:

$$\Pr[f(X_1, \dots, X_n) \leq (1 - \delta)\mu] = \Pr[e^{\lambda f(X_1, \dots, X_n)} \geq e^{\lambda(1 - \delta)\mu}] \leq \frac{\mathbf{E}[e^{\lambda f(X_1, \dots, X_n)}]}{e^{\lambda(1 - \delta)\mu}} \leq \frac{e^{(e^\lambda - 1)\mu}}{e^{\lambda(1 - \delta)\mu}}.$$

We choose $e^\lambda = 1 - \delta$ which yields

$$\Pr[f(X_1, \dots, X_n) \leq (1 - \delta)\mu] \leq \frac{e^{-\delta\mu}}{(1 - \delta)^{(1 - \delta)\mu}} \leq e^{-\mu\delta^2/2}$$

using $(1 - \delta)^{1-\delta} \geq e^{-\delta+\delta^2/2}$ for $\delta \in (0, 1]$. \square

D Lower-tail estimate for submodular functions under dependent rounding

In this section, we prove Theorem 1.4, i.e. an exponential estimate for the lower tail of the distribution of a monotone submodular function under randomized swap rounding. We note that the bound on the expected value of the rounded solution, $\mathbf{E}[f(R)] \geq \mu_0$, follows by the convexity of $F(x)$ along directions $\mathbf{e}_i - \mathbf{e}_j$ just like in [6]; we omit the details. The exponential tail bound is much more involved. We start by setting up some notation.

Notation. The rounding procedure starts from a convex linear combination of bases,

$$\mathbf{x}_0 = \sum_{i=1}^n \beta_i \mathbf{1}_{B_i}.$$

The rounding proceeds in stages, where in the first stage we merge the bases B_1, B_2 (randomly) into a new base C_2 , and replace $\beta_1 \mathbf{1}_{B_1} + \beta_2 \mathbf{1}_{B_2}$ in the linear combination by $\gamma_2 \mathbf{1}_{C_2}$, with $\gamma_2 = \beta_1 + \beta_2$. More generally, in the k -th stage, we merge C_k and B_{k+1} into a new base C_{k+1} (we set $C_1 = B_1$ in the first stage), and replace $\gamma_k \mathbf{1}_{C_k} + \beta_{k+1} \mathbf{1}_{B_{k+1}}$ in the linear combination by $\gamma_{k+1} \mathbf{1}_{C_{k+1}}$. Inductively, $\gamma_{k+1} = \gamma_k + \beta_{k+1} = \sum_{i=1}^{k+1} \beta_i$. After $n - 1$ stages, we obtain a linear combination $\gamma_n \mathbf{1}_{C_n}$ and $\gamma_n = \sum_{i=1}^n \beta_i = 1$; i.e., this is an integer solution.

We use the following notation to describe the vectors produced in the process:

- $\mathbf{b}_i = \beta_i \mathbf{1}_{B_i}$
- $\mathbf{c}_i = \gamma_i \mathbf{1}_{C_i}$
- $\mathbf{y}_k = \sum_{i=k}^n \mathbf{b}_i = \sum_{i=k}^n \beta_i \mathbf{1}_{B_i}$
- $\mathbf{x}_k = \mathbf{c}_{k+1} + \mathbf{y}_{k+2} = \gamma_{k+1} \mathbf{1}_{C_{k+1}} + \sum_{i=k+2}^n \beta_i \mathbf{1}_{B_i}$

In other words, \mathbf{b}_i are the initial vectors in the linear combination, which get gradually replaced by \mathbf{c}_i , and \mathbf{x}_k is the fractional solution after k stages.

We emphasize that \mathbf{x}_k denotes the entire fractional solution at a certain stage and not the value of its k -th coordinate. The coordinates of the fractional solution are the variables X_i . If we want to refer to the value of X_i after k stages, we use the notation $X_{i,k}$.

We work with the multilinear extension of a submodular function, $F(\mathbf{x}) = \mathbf{E}[f(\hat{\mathbf{x}})]$. In the following, we use the following shorthand notation and basic properties:

- $F_i(\mathbf{x})$ denotes the partial derivative $\frac{\partial F}{\partial X_i}$ evaluated at \mathbf{x} . The interpretation of $F_i(\mathbf{x})$ is the marginal value of i with respect to the fractional solution \mathbf{x} .
- We use $\mathbf{e}_i = \mathbf{1}_{\{i\}}$ to denote the canonical basis vector corresponding to element i .
- If only one variable is changing while others are fixed, $F(\mathbf{x})$ is a linear function. Therefore, we can use the following formula:

$$F(\mathbf{x} + t\mathbf{e}_i) = F(\mathbf{x}) + tF_i(\mathbf{x}).$$

- Due to submodularity, $\frac{\partial^2 F}{\partial X_i \partial X_j} \leq 0$ for any i, j . This implies that $F_i(\mathbf{x}) = \frac{\partial F}{\partial X_i}$ is non-increasing as a function of each coordinate of \mathbf{x} . If \mathbf{y} dominates \mathbf{x} in all coordinates ($\mathbf{x} \leq \mathbf{y}$), we have $F_i(\mathbf{x}) \geq F_i(\mathbf{y})$.

Proof overview. The random process in terms of the evolution of $F(x)$ is a submartingale, i.e. the value in each step can only increase in expectation. This is a good sign; however, a straightforward application of concentration bounds for martingales yields a dependency of the number of variables n which would render the bound meaningless. More refined bounds for martingales rely on bounds on the variance in successive steps. Unfortunately, these are also difficult to use since we do not have a good a priori bound on the variance in each step. The variance can depend on preceding steps and taking worst-case bounds leads to the same dependency on n as mentioned above.

In order to prove a bound which depends only on the parameters δ and μ_0 , we start from scratch and follow the standard recipe: estimate the exponential moment $\mathbf{E}[e^{\lambda(\mu_0 - f(R))}]$, where μ_0 is the initial value and R is the rounded solution. We decompose the expression $e^{\lambda(\mu_0 - f(R))}$ into a telescoping product:

$$e^{\lambda(\mu_0 - f(R))} = e^{\lambda(F(\mathbf{x}_0) - F(\mathbf{x}_{n-1}))} = e^{\lambda(F(\mathbf{x}_0) - F(\mathbf{x}_1))} \cdot e^{\lambda(F(\mathbf{x}_1) - F(\mathbf{x}_2))} \cdot \dots \cdot e^{\lambda(F(\mathbf{x}_{n-2}) - F(\mathbf{x}_{n-1}))}.$$

The factors in this product are not independent, but we can prove bounds on the conditional expectations $\mathbf{E}[e^{\lambda(F(\mathbf{x}_{k-1}) - F(\mathbf{x}_k))} \mid \mathbf{x}_0, \dots, \mathbf{x}_{k-1}]$, in other words conditioned on a given history of the rounding process. These bounds depend on the history, but we are able to charge the arising factors to the value of $\mu_0 = F(\mathbf{x}_0)$ in such a way that the final bound depends only on μ_0 .

We start from the bottom, by analyzing the basic rounding step for two variables. The following elementary inequality will be helpful.

Lemma D.1. *For any $p \in [0, 1]$ and $\xi \in [-1, 1]$,*

$$pe^{\xi(1-p)} + (1-p)e^{-\xi p} \leq e^{\xi^2 p(1-p)}.$$

Proof. If $\xi < 0$, we can replace ξ by $-\xi$ and p by $1 - p$; the statement of the lemma remains the same. So we can assume $\xi \in [0, 1]$.

Fix any $p \in [0, 1]$ and define $\phi_p(\xi) = e^{\xi^2 p(1-p)} - pe^{\xi(1-p)} - (1-p)e^{-\xi p}$. It is easy to see that $\phi_p(0) = 0$. Our goal is to prove that $\phi_p(\xi) \geq 0$ for $\xi \in [0, 1]$. Let us compute the derivative of $\phi_p(\xi)$ with respect to ξ :

$$\begin{aligned}\phi'_p(\xi) &= 2\xi p(1-p)e^{\xi^2 p(1-p)} - p(1-p)e^{\xi(1-p)} + p(1-p)e^{-\xi p} \\ &= p(1-p)e^{-\xi p} \left(2\xi e^{\xi^2 p(1-p)+\xi p} - e^\xi + 1 \right) \\ &\geq p(1-p)e^{-\xi p} \left(2\xi - e^\xi + 1 \right).\end{aligned}$$

For $\xi \in [0, 1]$, we have $e^\xi \leq 1 + 2\xi$ and hence $\phi'_p(\xi) \geq 0$. This means that $\phi_p(\xi)$ is non-decreasing and $\phi_p(\xi) \geq 0$ for $\xi \in [0, 1]$. \square

Note that the lemma does not hold for arbitrarily large ξ , e.g. when $p = 1/\xi^2$ and $\xi \rightarrow \infty$. Next, we apply this lemma to the basic step of the rounding procedure.

Lemma D.2. *Let $F(\mathbf{x})$ be the multilinear extension of a monotone submodular function with marginal values in $[0, 1]$, and let $\lambda \in [0, 1]$. Consider one elementary operation of randomized swap rounding, where two variables X_i, X_j are modified. Let \mathbf{x} denote the fractional solution before, \mathbf{x}' after this step, and let \mathcal{H} denote the complete history prior to this rounding step. Assume that the values of the two variables before the rounding step are $X_i = \gamma, X_j = \beta$. Then*

$$\mathbf{E}[e^{\lambda(F(\mathbf{x}) - F(\mathbf{x}'))} \mid \mathcal{H}] \leq e^{\lambda^2 \beta \gamma (F_j(\mathbf{x}) - F_i(\mathbf{x}))^2}$$

where $F_i(\mathbf{x}) = \frac{\partial F}{\partial X_i}(\mathbf{x})$ and $F_j(\mathbf{x}) = \frac{\partial F}{\partial X_j}(\mathbf{x})$.

Proof. Fix the history \mathcal{H} ; this includes the point \mathbf{x} before the rounding step. With probability $p = \frac{\gamma}{\beta+\gamma}$, the rounding step is $X'_i = X_i + \beta$ and $X'_j = X_j - \beta$. I.e., $\mathbf{x}' = \mathbf{x} + \beta \mathbf{e}_i - \beta \mathbf{e}_j$. Since $F(\mathbf{x})$ is linear when only one coordinate is modified, we get

$$F(\mathbf{x}') = F(\mathbf{x}) + \beta F_i(\mathbf{x}) - \beta F_j(\mathbf{x} + \beta \mathbf{e}_i).$$

By submodularity, $F_j(\mathbf{x} + \beta \mathbf{e}_i) \leq F_j(\mathbf{x})$ and hence

$$F(\mathbf{x}') = F(\mathbf{x}) + \beta F_i(\mathbf{x}) - \beta F_j(\mathbf{x} + \beta \mathbf{e}_i) \geq F(\mathbf{x}) + \beta(F_i(\mathbf{x}) - F_j(\mathbf{x})).$$

With probability $1 - p$, we set $X'_i = X_i - \gamma$ and $X'_j = X_j + \gamma$. By similar reasoning, in this case we get

$$F(\mathbf{x}') = F(\mathbf{x}) - \gamma F_i(\mathbf{x}) + \gamma F_j(\mathbf{x} - \gamma \mathbf{e}_i) \geq F(\mathbf{x}) - \gamma(F_i(\mathbf{x}) - F_j(\mathbf{x})).$$

Taking expectation over the two cases, we get

$$\begin{aligned}\mathbf{E}[e^{\lambda(F(\mathbf{x}) - F(\mathbf{x}'))} \mid \mathcal{H}] &\leq pe^{\lambda \beta (F_j(\mathbf{x}) - F_i(\mathbf{x}))} + (1-p)e^{-\lambda \gamma (F_j(\mathbf{x}) - F_i(\mathbf{x}))} \\ &= pe^{\lambda(1-p)(\beta+\gamma)(F_j(\mathbf{x}) - F_i(\mathbf{x}))} + (1-p)e^{-\lambda p(\beta+\gamma)(F_j(\mathbf{x}) - F_i(\mathbf{x}))}.\end{aligned}$$

We invoke Lemma D.1 with $\xi = \lambda(\beta + \gamma)(F_j(\mathbf{x}) - F_i(\mathbf{x}))$ (we have $|\xi| \leq 1$ due to $\lambda, \beta + \gamma, F_i(\mathbf{x}), F_j(\mathbf{x})$ all being in $[0, 1]$). We get

$$\mathbf{E}[e^{\lambda(F(\mathbf{x}) - F(\mathbf{x}'))} \mid \mathcal{H}] \leq e^{\xi^2 p(1-p)} = e^{\lambda^2 \beta \gamma (F_j(\mathbf{x}) - F_i(\mathbf{x}))^2}.$$

\square

Note that the exponent on the right-hand side of Lemma D.2 corresponds to the variance in one step of the rounding procedure. The next lemma estimates these contributions, aggregated over one stage of the rounding process, i.e., the merging of the bases C_k and B_{k+1} . The exponent on the right-hand side of Lemma D.3 corresponds to the variance of the random process accumulated over the k -th stage. It is crucial that we compare this quantity to certain values which can be eventually charged to μ_0 .

Lemma D.3. *Let $F(\mathbf{x})$ be the multilinear extension of a monotone submodular function with marginal values in $[0, 1]$, and let $\lambda \in [0, 1]$. Consider the k -th stage of the rounding process, when bases C_k and B_{k+1} (with coefficients γ_k and β_{k+1}) are merged into C_{k+1} . The fractional solution before this stage is \mathbf{x}_{k-1} and after this stage \mathbf{x}_k . Conditioned on any history \mathcal{H} of the rounding process throughout the first $k - 1$ stages,*

$$\mathbf{E}[e^{\lambda(F(\mathbf{x}_{k-1}) - F(\mathbf{x}_k))} \mid \mathcal{H}] \leq e^{\lambda^2(\beta_{k+1}F(\mathbf{c}_k) + \gamma_k(F(\mathbf{y}_{k+1}) - F(\mathbf{y}_{k+2})))}.$$

Proof. The k -th stage merges bases C_k and B_{k+1} into C_{k+1} by taking elements in pairs and performing rounding steps as in Lemma D.2. Let us denote the pairs of elements considered by the rounding procedure $(c_1, b_1), \dots, (c_d, b_d)$, where $C_k = \{c_1, \dots, c_d\}$ and $B_{k+1} = \{b_1, \dots, b_d\}$. The matching is not determined beforehand: (c_2, b_2) might depend on the random choice between c_1, b_1 , etc. In the following, we drop the index k and denote by \mathbf{x}^i the fractional solution obtained after processing $(c_1, b_1), \dots, (c_i, b_i)$. We start with $\mathbf{x}^0 = \mathbf{x}_{k-1}$ and after processing all d pairs, we get $\mathbf{x}^d = \mathbf{x}_k$. We also replace β_{k+1}, γ_k simply by β, γ . We denote by \mathcal{H}_i the complete history prior to the rounding step involving (c_{i+1}, b_{i+1}) ; in particular, this includes the fractional solution \mathbf{x}^i .

Using Lemma D.2 for the rounding step involving (c_{i+1}, b_{i+1}) , we get

$$\mathbf{E}[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^{i+1}))} \mid \mathcal{H}_i] \leq e^{\lambda^2\gamma\beta(F_{c_{i+1}}(\mathbf{x}^i) - F_{b_{i+1}}(\mathbf{x}^i))^2} \leq e^{\lambda^2\gamma\beta(F_{c_{i+1}}(\mathbf{x}^i) + F_{b_{i+1}}(\mathbf{x}^i))},$$

using the fact that the partial derivatives $F_j(\mathbf{x}^i)$ are in $[0, 1]$.

Further, we modify the exponent of the right-hand side as follows. The vector \mathbf{x}^i is obtained after processing i pairs and still contains the coordinates c_{i+1}, \dots, c_d of $\mathbf{c}_k = \gamma \mathbf{1}_{C_k}$ untouched: in other words, $\mathbf{x}^i \geq \gamma \mathbf{1}_{\{c_{i+1}, \dots, c_d\}}$. Let us define

- $\mathbf{c}^i = \gamma \mathbf{1}_{\{c_{i+1}, \dots, c_d\}}$.

I.e., $\mathbf{x}^i \geq \mathbf{c}^i \geq \mathbf{c}^{i+1}$. By submodularity, we have $F_{c_{i+1}}(\mathbf{x}^i) \leq F_{c_{i+1}}(\mathbf{c}^{i+1})$.

Similarly, the vector \mathbf{x}^i also contains the coordinates b_{i+1}, \dots, b_d of \mathbf{b}_{k+1} and all of $\mathbf{y}_{k+2} = \sum_{j=k+2}^n \mathbf{b}_j$ unchanged: $\mathbf{x}^i \geq \beta \mathbf{1}_{\{b_{i+1}, \dots, b_d\}} + \mathbf{y}_{k+2}$. Let us define

- $\mathbf{y}^i = \beta \mathbf{1}_{\{b_{i+1}, \dots, b_d\}} + \mathbf{y}_{k+2}$.

I.e., $\mathbf{x}^i \geq \mathbf{y}^i \geq \mathbf{y}^{i+1}$. By submodularity, we get $F_{b_{i+1}}(\mathbf{x}^i) \leq F_{b_{i+1}}(\mathbf{y}^{i+1})$. Therefore, we can write

$$\mathbf{E}[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^{i+1}))} \mid \mathcal{H}_i] \leq e^{\lambda^2\gamma\beta(F_{c_{i+1}}(\mathbf{c}^{i+1}) + F_{b_{i+1}}(\mathbf{y}^{i+1}))}. \quad (5)$$

We claim that by induction on $d - i$, this implies

$$\mathbf{E}[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^d))} \mid \mathcal{H}_i] \leq e^{\lambda^2(\beta F(\mathbf{c}^i) + \gamma(F(\mathbf{y}^i) - F(\mathbf{y}^d)))} \quad (6)$$

for all $i = 0, \dots, d$. For $i = d$, the claim is trivial. For $i < d$, we can write

$$\mathbf{E}[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^d))} \mid \mathcal{H}_i] = \mathbf{E}\left[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^{i+1}))} \mathbf{E}[e^{\lambda(F(\mathbf{x}^{i+1}) - F(\mathbf{x}^d))} \mid \mathcal{H}_{i+1}] \mid \mathcal{H}_i\right]$$

and using the inductive hypothesis (6) for $i + 1$,

$$\begin{aligned} \mathbf{E}[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^d))} \mid \mathcal{H}_i] &\leq \mathbf{E}\left[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^{i+1}))} \cdot e^{\lambda^2(\beta F(\mathbf{c}^{i+1}) + \gamma(F(\mathbf{y}^{i+1}) - F(\mathbf{y}^d)))} \mid \mathcal{H}_i\right] \\ &= e^{\lambda^2(\beta F(\mathbf{c}^{i+1}) + \gamma(F(\mathbf{y}^{i+1}) - F(\mathbf{y}^d)))} \cdot \mathbf{E}\left[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^{i+1}))} \mid \mathcal{H}_i\right] \end{aligned}$$

where we used the fact that the inductive bound is determined by \mathcal{H}_i , and so we can take it out of the expectation (it depends only on the sets $\{c_{i+2}, \dots, c_d\}$ and $\{b_{i+2}, \dots, b_d\}$ which are determined even before performing the rounding step on (c_{i+1}, b_{i+1})). Taking logs and using (5) to estimate the last expectation, we obtain

$$\begin{aligned} & \log \mathbf{E}[e^{\lambda(F(\mathbf{x}^i) - F(\mathbf{x}^d))} \mid \mathcal{H}_i] \\ & \leq \lambda^2 \left(\beta F(\mathbf{c}^{i+1}) + \gamma(F(\mathbf{y}^{i+1}) - F(\mathbf{y}^d)) \right) + \lambda^2 \gamma \beta \left(F_{c_{i+1}}(\mathbf{c}^{i+1}) + F_{b_{i+1}}(\mathbf{y}^{i+1}) \right) \\ & = \lambda^2 \left(\beta(F(\mathbf{c}^{i+1}) + \gamma F_{c_{i+1}}(\mathbf{c}^{i+1})) + \gamma(F(\mathbf{y}^{i+1}) + \beta F_{b_{i+1}}(\mathbf{y}^{i+1}) - F(\mathbf{y}^d)) \right) \\ & = \lambda^2 \left(\beta F(\mathbf{c}^i) + \gamma(F(\mathbf{y}^i) - F(\mathbf{y}^d)) \right) \end{aligned}$$

where we used $F(\mathbf{c}^{i+1}) + \gamma F_{c_{i+1}}(\mathbf{c}^{i+1}) = F(\mathbf{c}^i)$ and $F(\mathbf{y}^{i+1}) + \beta F_{b_{i+1}}(\mathbf{y}^{i+1}) = F(\mathbf{y}^i)$ (see the definitions of $\mathbf{c}^i, \mathbf{y}^i$ above).

This proves our inductive claim (6). For $i = 0$, since $\mathbf{x}^0 = \mathbf{x}_{k-1}$, $\mathbf{x}^d = \mathbf{x}_k$, $\mathbf{c}^0 = \mathbf{c}_k$, $\mathbf{y}^0 = \mathbf{y}_{k+1}$ and $\mathbf{y}^d = \mathbf{y}_{k+2}$, this gives the statement of the lemma. \square

Now we can proceed finally to the proof of Theorem 1.4.

Proof. We prove inductively the following statement: For any k and any $\lambda \in [0, 1]$,

$$\mathbf{E}[e^{\lambda(\mu_0 - F(\mathbf{x}_k))}] \leq e^{\lambda^2(\mu_0(1 + \sum_{i=1}^k \beta_{i+1}) - F(\mathbf{y}_{k+2}))}. \quad (7)$$

We remind the reader that $\mu_0 = F(\mathbf{x}_0)$, \mathbf{x}_k is the fractional solution after k stages, and $\mathbf{y}_{k+2} = \sum_{i=k+2}^n \mathbf{b}_i$. We proceed by induction on k .

For $k = 0$, the claim is trivial, since $F(\mathbf{y}_2) \leq F(\mathbf{x}_0) = \mu_0$ by monotonicity. For $k \geq 1$, we unroll the expectation as follows:

$$\mathbf{E}[e^{\lambda(\mu_0 - F(\mathbf{x}_k))}] = \mathbf{E}\left[e^{\lambda(\mu_0 - F(\mathbf{x}_{k-1}))} \mathbf{E}[e^{\lambda(F(\mathbf{x}_{k-1}) - F(\mathbf{x}_k))} \mid \mathcal{H}]\right]$$

where \mathcal{H} is the complete history prior to stage k (up to \mathbf{x}_{k-1}). We estimate the inside expectation using Lemma D.3:

$$\mathbf{E}[e^{\lambda(F(\mathbf{x}_{k-1}) - F(\mathbf{x}_k))} \mid \mathcal{H}] \leq e^{\lambda^2(\beta_{k+1}F(\mathbf{c}_k) + \gamma_k(F(\mathbf{y}_{k+1}) - F(\mathbf{y}_{k+2})))} \leq e^{\lambda^2(\beta_{k+1}F(\mathbf{x}_{k-1}) + F(\mathbf{y}_{k+1}) - F(\mathbf{y}_{k+2}))}$$

using monotonicity, $\mathbf{c}_k \leq \mathbf{x}_{k-1}$, $\mathbf{y}_{k+2} \leq \mathbf{y}_{k+1}$ and $\gamma_k \leq 1$. Therefore,

$$\begin{aligned} \mathbf{E}[e^{\lambda(\mu_0 - F(\mathbf{x}_k))}] & \leq \mathbf{E}\left[e^{\lambda(\mu_0 - F(\mathbf{x}_{k-1}))} e^{\lambda^2(\beta_{k+1}F(\mathbf{x}_{k-1}) + F(\mathbf{y}_{k+1}) - F(\mathbf{y}_{k+2}))}\right] \\ & = e^{\lambda^2(\beta_{k+1}\mu_0 + F(\mathbf{y}_{k+1}) - F(\mathbf{y}_{k+2}))} \mathbf{E}\left[e^{(\lambda - \lambda^2\beta_{k+1})(\mu_0 - F(\mathbf{x}_{k-1}))}\right]. \end{aligned}$$

By the inductive hypothesis (7) with $\lambda' = \lambda - \lambda^2\beta_{k+1} \in [0, 1]$,

$$\mathbf{E}[e^{(\lambda - \lambda^2\beta_{k+1})(\mu_0 - F(\mathbf{x}_{k-1}))}] \leq e^{(\lambda - \lambda^2\beta_{k+1})^2(\mu_0(1 + \sum_{i=1}^{k-1} \beta_{i+1}) - F(\mathbf{y}_{k+1}))} \leq e^{\lambda^2(\mu_0(1 + \sum_{i=1}^{k-1} \beta_{i+1}) - F(\mathbf{y}_{k+1}))}.$$

In the last inequality we used $F(\mathbf{y}_{k+1}) \leq \mu_0$, which holds by monotonicity. Plugging this into the preceding equation,

$$\begin{aligned} \mathbf{E}[e^{\lambda(\mu_0 - F(\mathbf{x}_k))}] & \leq e^{\lambda^2(\beta_{k+1}\mu_0 + F(\mathbf{y}_{k+1}) - F(\mathbf{y}_{k+2}))} e^{\lambda^2(\mu_0(1 + \sum_{i=1}^{k-1} \beta_{i+1}) - F(\mathbf{y}_{k+1}))} \\ & = e^{\lambda^2(\mu_0(1 + \sum_{i=1}^k \beta_{i+1}) - F(\mathbf{y}_{k+2}))} \end{aligned}$$

which proves (7). Finally, for $k = n - 1$ we obtain $F(\mathbf{x}_{n-1}) = f(R)$ where R is the rounded solution, $\mathbf{y}_{n+1} = 0$, and

$$\mathbf{E}[e^{\lambda(\mu_0 - f(R))}] \leq e^{\lambda^2\mu_0(1 + \sum_{i=1}^{n-1} \beta_{i+1})} \leq e^{2\lambda^2\mu_0} \quad (8)$$

because $\sum_{i=1}^{n-1} \beta_{i+1} \leq 1$. The final step is to apply Markov's inequality to the exponential moment. From Markov's inequality and Equation (8), we get

$$\Pr[f(R) \leq (1 - \delta)\mu_0] = \Pr\left[e^{\lambda(\mu_0 - f(R))} \geq e^{\lambda\delta\mu_0}\right] \leq \frac{\mathbf{E}[e^{\lambda(\mu_0 - f(R))}]}{e^{\lambda\delta\mu_0}} \leq e^{2\lambda^2\mu_0 - \lambda\delta\mu_0}.$$

A choice of $\lambda = \delta/4$ gives the statement of the theorem. \square